

# 机械设计手册

新版

6



MACHINE DESIGN HANDBOOK



机械工业出版社  
CHINA MACHINE PRESS

## 权威

国内机械工业知名学者和机械设计专家研究并执笔编写，保证了本手册的高质量和权威性。

全面反映国内外机械设计的最新成果，所涉及的设计方法和理论与国际接轨，反映国内外设计水平。技术数据、产品数据准确可靠。

## 系统

从设计理念、设计方法到常用数据，系统地总结了机械设计各专业的技术内容。将新思维、新方法与设计实践融会贯通到机械设计的全过程中。

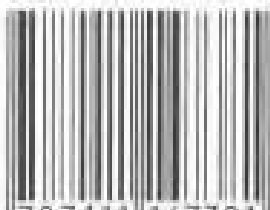
## 实用

为机械设计实践提供了全面的技术资料，包括机械、电器、液压、气动等专业。给出了机械设计基础资料，机械设计常用材料，机械设计方法，常用零部件的类型、规格、尺寸、设计要点，常用结构等具体内容。为各种机械产品的设计工作全面提供计算方法、主要技术参数、选型原则、典型结构和设计计算实例。全部采用2004年6月止的现行国际、国家及行业最新标准。实现了信息充分、数据全面、取材广泛、结构多样、产品新颖，并通过合理编排，力求便于查阅，使用方便。

## 先进

提供了当今国际、国内公认的先进设计理论、方法及数据资料，技术前瞻与国际设计水平同步。

ISBN 7-111-14738-3



9 787111 147381 >

定价：102.00元  
中国标准出版社  
地址：北京机械工业出版社  
电话：010-68311636  
网址：http://www.csp.com.cn  
E-mail: csp@bjcsp.com.cn



# 机械 设计 手册

新 版

第 6 卷

机械设计手册编委会



机械工业出版社

本书是在前几版的基础上,吸收了近年来新的设计方法及最新国家标准,全面、系统地介绍了所有现代设计和常规设计方法、数据、图表,内容丰富,具有信息量大,标准新、取材广、规格全、常用结构多、并增加了许多国内外常用的新产品的结构、规格、选用范围,实用性强、查找方便等特点。

全书共分常用资料,机械零部件与传动设计(一)、(二),液压、气动、液力传动与控制,机械设计基础,现代设计方法及应用等6卷50篇。

本卷主要介绍机械设计领域中创新、绿色产品、智能、并行、有限元、虚拟、快响应变型、模块化、优化、可靠性、实验、反求、互联网上合作等现代设计方法及其应用技术。并对各种设计方法举例说明,书中最后还介绍了机械设计中现代设计主流软件的应用范围等。

## 图书在版编目(CIP)数据

机械设计手册.第6卷/机械设计手册编委会编著.—3版.—北京:机械工业出版社,2004.8

ISBN 7-111-14738-3

I. 机... II. 机... III. 机械设计-技术手册 N. TH122-62

中国版本图书馆CIP数据核字(2004)第068533号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

责任编辑:曲彩云 李建秀 张亚秋 版式设计:张世琴

责任校对:张媛 封面设计:姚毅 责任印制:李妍

北京蓝海印刷有限公司印刷·新华书店北京发行所发行

2004年8月第3版·第1次印刷

787mm×1092mm<sup>1</sup>/<sub>16</sub>·76.75印张·3插页·2663千字

0 001—6 000册

定价:102.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线电话(010)68993821、88379646

封面无防伪标均为盗版

## 机械设计手册编委会

主任：王文斌

副主任：林忠钦 严隽琪 李奇 谢里阳 汪恺 孙慧波

委员：鄂中凯 崔虹曼 方昆凡 周康年 吴宗泽 樊文萱 黄万吉

吴自通 徐秀彦 徐鹏 朱孝录 施永乐 王起龙 巩云鹏

李立行 程乃士 王德俊 李元科 卜炎 施高义 郑洪生

周恩涛 宗跃 唐恒龄 高敏 何德方 欧宗瑛 黄雨华

郭宝柱 张健民 史家顺 陈铭 蔡建国 王安麟 钟廷修

蒋寿伟 王石刚 邹慧君 金焯 谢友柏 蒋祖华 曲彩云

# 新版前言

《机械设计手册》自出版以来，在机械设计实践中发挥了重要的基础性作用，先后荣获全国优秀科技图书二等奖、机械电子工业部科技进步二等奖，是全国优秀科技畅销书，在社会上有较高的知名度，影响广泛，深得广大工程技术人员的厚爱。

机械设计是机械工业的基础技术。科技成果要转化为有竞争力的新产品，设计起着关键性作用。设计工作的质量和水平，直接关系到产品质量、性能和技术经济效益。特别是在产品创新和创造方面，工业发达国家都极为重视机械设计工作，不断地研制出适应市场需要的机电产品。近几年来，由于科学技术不断发展，加之我国加入世贸组织以后国际技术交流更直接、便捷，使我国在机械设计领域有了长足的进步，取得了丰硕的成果，发现了许多新材料、新的设计理念和设计方法，这些都需要我们及时归纳总结，全面准确地提供给广大读者。为使《机械设计手册》紧跟时代步伐，满足广大读者需要，我们组织了这次《机械设计手册》的新版修订工作。

这次修订，根据广大工程技术人员实际需要和阅读习惯，在保持手册先进性、系统性的前提下，我们调整了卷、篇、章的框架结构，重新编排，并补充了机械设计应用方面的内容。更加突出实用便查，让技术人员感到既有很多成熟使用的现代设计方法，又能更方便、快捷地查到所需内容。

这次修订取材仍根据基本、常用、关键、发展的原则，强调准确性。我们认真细致地对各种数据、图表等进行分析、核对和验证，对一些局部性的技术经验和测试结果，为了做到准确、可靠，作者亲自南下北上，取得第一手试验数据资料，以确定选取范围，做到了精益求精。对国外资料，以常用和新材料为主，主要从选用角度编写，重点介绍材料的技术参数、性能特点、适用范围和应用技术等，为如何正确选择和合理使用提供依据，以发挥其最佳性能和经济效果，这些都是作者经过长期生产实践积累的宝贵资料。本书数据、资料全部来自国家最新标准、规范和其他权威机构，设计方法、公式选择、参数选取原则均经过长期实践检验，典型结构设计、计算实例均来自工程实践。为了突出反映先进性，增加了许多适合我国国情的新材料、新方法、新技术、新工艺，广泛收集具有先进水平的新产品。

重点修订内容：

## (1) 更换补充新标准

书中所涉及的标准均采用现行国际、国家及行业最新标准，这对提高我国机械设计标准化水平，促进机械产品走向世界都将起到促进作用。同时，对一些新旧标准过渡有难点的内容做了详细指导，如渐开线圆柱齿轮部分引入最新国家标准 GB/T10095.1~2—2001 渐开线圆柱齿轮精度和 GB/Z18620.1~4—2002 圆柱齿轮检验实施规范；考虑到新旧标准应用的过渡期，保留了渐开线圆柱齿轮精度标准 GB/T100095--1988 的相关内容，给出了新旧标准精度对照表及旧标准向新标准的转化方法，同时给出了新旧标准标注的齿轮工作图示例，以方便读者使用。

锥齿轮强度计算采用最新颁布的国家标准 GB/T10062.1~3--2003，为便于采用新标准，增加了强度计算用到的锥齿轮的端面当量齿轮和法面当量齿轮参数计算公式，给出了用新标准计算的两个锥齿轮传动设计算例。把现有国内资料中关于锥齿轮几何设计中的“参考点”改为“基

准点”，以忠实原意。

在轴承中增加了 2004 年 6 月发布的滚动轴承代号方法的补充规定，以作为轴承标准的补充内容。

#### (2) 增加新材料、新结构

由于新材料、新结构不断出现，故在这次修订中，均适当纳入新内容，更换老内容，使手册更新、更适用。

如由于工程塑料和复合材料的力学性能有了很大的提高，又具有价廉、防腐、防锈蚀等一系列优点，故其应用发展很快，我们加强了这方面的介绍，书中还大量增加了新的结构类型和应用实例。

在轴承部分增加了带座外球面轴承、专用轴承、直线运动滚动支撑、关节轴承，以及国外轴承、钢球、钢种、润滑油等的代号和国内外对照表。

#### (3) 介绍新产品，删掉淘汰产品

机械设计中经常选用的一些基础产品发生了很大的变化，现在产品的分类、代号方法、设计计算等内容完全与国际接轨，与老产品相比完全不同。对于这些内容，我们必须加以修正，以适应我国机械产品进入国际市场的需要。因而，我们删掉了部分淘汰产品内容，并将国内主要厂家新开发推出的具有较大影响的新产品进行了补充，并适当增加了国内选用较多的国外产品，如为适应目前国内市场上进口液压、液力、气动元器件所占份额逐渐增大的现状，特增加了大篇幅的内容，主要推出世界著名的几大品牌应用较广泛的产品，以满足专业技术人员的需要。

近年来我国出现的新的减速器或形成了新标准的减速器，都在不断地向国外的新技术看齐，也收入本手册中供大家参考。此外，还及时引入了国外的产品，如 SEW 和 PIV 的产品，一则是为了向国际靠拢（接轨），二则是让设计人员方便选用国外知名公司的适用产品。

#### (4) 增加了反映现代科技的新内容

在“机械振动的控制”一章内，不仅填补了对冲击、对随机振动的隔离等国内外空白内容，还重点增加了利用振动信号的测量、分析、反馈及跟踪等先进技术，进行最优控制、自适应控制、预测控制、模糊控制等振动的主动控制的新内容。增加了“振动的利用”一章，使读者不仅能查到对有害振动进行控制的内容，还能查到振动的有益内容，包括振动利用的方法、步骤、设计与计算，利用振动原理工作的各类机械。

为适应现代机械设计的要求，增加了“模态分析与参数识别”一章。在机械设计中应用此章的内容，能对机械给定的动力特性，识别出机械应有的物理参数，或者已知机械的物理参数，识别出其动力特性。还能识别出很难准确计算和直接测量的机械动载荷，更能在机械运行时“在线”识别其动态特性。这些都将提高机械动态优化设计的功能与效率。

#### (5) 充实增加了现代设计方法的应用技术

近年来，机械设计的理论和方法在国内外取得了很大的发展，我国设计人员要在产品设计方面赶超国际水平，必须掌握先进的设计理论和方法。由于机械设计涉及面广，即使是经验丰富的设计师，也难以及时、全面地掌握这些理论与方法，所以本书重点介绍了当今发达工业国家流行的成熟的设计理论与方法，增强广大工程技术人员的创新意识和能力。如通过对有限元设计、创新设计、虚拟设计、优化设计、并行设计、智能设计、机电一体化设计及其相应工具软件等内容的详细阐述，使得工程技术人员在日常设计工作中融入现代设计的理论与方法，并注重相应流行

软件的应用,达到提高设计水平和设计效率的目的。如有限元设计部分,详细介绍将“有限元设计”如何应用于实际设计工作中,以适应国际流行的针对重大机电产品(包括新产品)研发及销售过程中需配备有限元分析软件的需要。又如为加强环保,进行绿色设计,生产绿色产品,在各国均受到普遍欢迎,工业发达国家更是十分重视环保这方面的工作,所以本手册也介绍了如何实施机电产品的绿色设计。所有这些内容,都是现代设计工作者必不可少的最新的必备知识,而其他同类工具书却极少涉及这些内容。

可靠性、优化、疲劳强度、摩擦学等理论的新发展,可以为机械设计人员提供当代的先进资料,其中有些是最新的研究成果。设计人员可以较方便地找到自己需要的方法或解决问题的线索,并为进一步深入开展工作打下良好的基础。

本书作者都是具有丰富的设计知识和技能,具有出色的机械设计实践和研究经验的本学科知名学者和机械设计专家,他们统观全局、采实撷精、为本书修订奠定了可靠的保证。

## 第 1 版前言

《机械设计手册》是继《机械工程手册》之后出版的一部大型机械设计专业技术工具书。

机械工业担负着向国民经济各部门,包括工业、农业和社会生活各个方面提供各种性能先进、价格低廉、使用安全可靠的技术装备的任务,所以在现代化建设中是举足轻重的。市场竞争的生命力在于产品的水平。任何科技成果要转变为有竞争力的商品,设计起着关键性的作用。机械设计是机械产品研制的第一道工序,设计工作的质量和水平,直接关系到产品质量、性能、研制周期和技术经济效益。工业发达国家都十分重视产品设计:日本认为,工业发达是企业对产品设计高度重视的结果;美国认为,设计是一本万利的事,对产品设计投资 1 美元,带来的利润却是 1500 美元;英国认为,产品设计是英国工业的命脉,英国工业革新必须以设计为中心,始终应把产品设计作为企业的头等大事,应时常探索研究使产品设计尽善尽美;法国认为,设计是工业的生命,要培养超一流设计大师,要大胆启用有才华有实践工作经验的设计人员。

这里,有必要回顾一下机械和机械设计发展的历史。机械的发明和发展,是先由几种简单工具开始的。石器时代的石刀、石斧,只是为了能省力或便于用力。后来发展到利用杠杆原理制作灌溉或扬水用的桔槔,利用滑轮原理制作重物提升用的辘轳等简单机械。这些机械所需的原动力是直接出自人的本身。为了省力和扩大力,开始时利用牲畜力,后来利用风力和火力。待到 18 世纪 60 年代发明了蒸汽机,作为动力带动了纺织机、磨粉机、鼓风机、工作母机和铁路机车,促进了冶金、轮船和火车等工业的发展。到 19 世纪 60 年代,出现了第一台直流发电机,到 19 世纪 80 年代,研制成功了交流发电机和交流电动机,20 世纪初,电动机已在工业生产中取代了蒸汽机,成为驱动各种工作机械的基本动力。电气技术的应用,使机械工业得到了高速的发展。工业的发展,要求围绕机械设计制造的基础理论和设计方法,能适应当时机械工业的形势。到 18~19 世纪,材料科学、结构力学、弹性力学、流体力学、热力学、制图和公差等,都分别发展成为一门独立学科。但由于机械设计的复杂性,还需将这些学科在应用于设计时作某些简化假设,再加上设计人员的经验,逐渐形成了一整套机械设计方法。在这套设计方法中,要应用一些经验设计方法、经验设计公式和经验系数等,称之为常规设计或传统设计。

1946 年世界上第一台电子数字计算机诞生。经历了电子管、半导体、集成电路和大型集成电路的发展,电子计算机在机械设计中已广为采用。电子计算机的发展,使有限元法、优化设计和计算机辅助设计等成为可能。加上材料科学、计算力学、摩擦学和设计理论等的发展,逐渐形成了一套现代设计理论和方法。现代设计的特点为:(1)从静态设计到动态设计;(2)从单项设计指标到综合设计指标;(3)从常规设计到精确设计;(4)从手算设计到广泛应用计算机的设计。常规设计是不可缺少的,但对于培养具有更广阔视野的设计人员来说显得非常不够。近二、三十年,设计方法更为科学化、系统化、完善化和现代化了,虽然如此,常规设计仍然是重要基础。

由于机械产品品种繁多,除一些重要的机械产品(如机床等)有专业手册,加上综合性的《机械工程手册》外,编写一部能统贯整个机械设计领域,主要写机械设计共性内容,具有现代设计水平,实用性强,为机械设计学科领域的机械设计人员、科研和教学工作者查阅使用的《机械设计手册》,实属当务之急。为此,机械工业出版社于 1985 年冬着手组织全国专家、学者进行《机械设计手册》的编写工作。

本手册是在现代设计方法在我国经历了宣传普及阶段并在设计中初步取得成果、新的设计标准规范陆续制订公布的有利时机完成编写工作的。在制订编写提纲过程中,广泛听取了各方面的意见,将设计作为一个整体来考虑,不仅要考虑强度和润滑等常规设计注意的问题,还要考虑便于制造、技术经济指标合理和美观等方面,贯彻“四性”(实用性、整体性、科学性、先进性)精神,立足于80年代机械设计水平进行编写。手册中的计量单位一律采用国家法定计量单位,原有的数据单位,还没换成法定单位的,我们一律换算成法定单位。标准均为现行标准。

本手册共有42篇,分5卷出版。第1卷共7篇。第1篇机械设计总论,对机械设计的地位、设计遵循原则、设计的内容和设计方法作战略性的描述,使读者对机械设计有整体性理解。后面6篇是机械设计的基础理论和基本数据,各篇尽量用较小的篇幅写出覆盖面广的现代设计所需的实用内容。第2卷共10篇,是现代设计理论和设计方法,其中第8篇机构及机械系统设计,是机械设计的第一步骤,它是方案设计的主要内容。考虑到现代设计中的计算机应用,故以数值解法为主,代替了传统的图解法。第9篇造型设计和人机工程,介绍机械设计中如何考虑机器的形体和色彩,如何考虑操作者的人体尺寸、出力大小和视力范围等。第10篇价值工程,介绍机械设计中技术经济指标的计算以及评价和决策。下面几篇疲劳强度设计、蠕变设计、可靠性设计、优化设计、计算机辅助设计,都是一些现代设计方法。第16篇是计算机辅助设计所用的“数据库”,第15篇是与现代设计密切相关的“失效分析和故障诊断”。这些篇大多是现有手册中没有的,个别篇虽然少数手册中有类似的篇名,但本手册是从现代设计的要求出发进行编写,内容新而且深入。第3卷共8篇,第4卷共11篇,是机械零部件设计部分。虽然其中大部分篇名,在一些机械设计手册中也有,但本手册采用了最新的标准规范,尽量与现代设计相结合,所以各篇中都有一些内容,甚至整篇内容在一般手册中是没有的。一些重要的设计计算,另备有设计软件包。第5卷共6篇,是各种传动、机械自动化和工业机器人。其中工业机器人是机电仪一体化的典型产品,作为本手册的终篇,对贯彻本手册编写意图是有深刻含意的。为与本手册精神相一致,工业机器人也只写其共性部分。

《机械设计手册》是一部千万字的巨著,参加编写人员近200名,组织工作繁重。为了统一编写精神,经多次讨论确定了编写体例,按篇确定主编,由主编提出编写人员,召开编写会,审查各篇的编写提纲,按篇确定2~4位审稿人,初稿完成后送审,审稿意见与编写人见面,共同商量改稿意见,在此基础上,部分篇召开了审稿会。待到条件成熟,按卷召开定稿会。所以,本手册的出版,是在国内大专院校、科研院所和工厂的教授、研究人员和工程师的积极参加下完成的,并得到机械工业出版社、机械电子工业部科技司和东北工学院等单位的大力支持,这是本手册能够在较短的时间内从组织编写到出书的有力保证,在此谨向大家表示诚挚的感谢,并衷心希望广大读者提出批评意见,使本手册在修订时能有所改进。

徐 灏

1988年11月



## 第2版前言

《机械设计手册》自1990年出版至今已有10年,曾8次印刷,销售10万多套,得到了广大读者的关心、支持和好评,获第七届全国优秀科技图书二等奖,原机械电子工业部科技进步二等奖。

现在,《机械设计手册》(第2版)又与广大读者见面了!

2版修订是在1版的基础上,调整结构、更新内容、完善不足、更新标准、突出实用,让广大机械设计人员更方便快捷地查到所需内容。

### 一、修订的重点

1. 充实和更新技术内容。在重点反映国内外机械设计领域的新技术、新材料的同时,加强了自动化技术、计算机技术等在设计中的应用。现代设计方法和应用等都增设了新篇章。对于一些有发展前景的新设计方法,也作了相应介绍。2版新增设了电动机和常用低压电器、创新设计、绿色产品设计、并行设计、虚拟设计、快速响应变型设计和反求设计、机电一体化系统设计、现代设计主流软件、零部件设计常用基础标准、传动总论等10篇;重新编写了摩擦学设计、优化设计、计算机辅助设计、带传动和链传动、齿轮传动、滚动轴承、滑动轴承、气压传动与控制、液压传动与控制等9篇。其他各篇也作了较大程度的修改或更新。

2. 突出重点,务求实用。在总体结构和内容设置上作了一定调整,精简了基础理论部分内容,注意收集设计实践的经验和数据,使手册结构更趋合理,内容更切实际,更方便查阅。

3. 更换最新标准。根据到2000年6月为止颁发的国家或行业现行标准及技术规范,重新更换了旧的标准,体现了技术内容和数据的可靠性。

### 二、内容和结构

2版主要包括常用资料和设计基础、现代设计方法及应用、机械零部件设计、机械传动设计、流体传动与控制等部分,共44篇,分为5卷。

1. 常用资料和设计基础 机械设计总论、常用资料和数学公式、机械工程材料、机械设计力学基础、实验应力分析、机械振动和噪声、造型设计和人机工程、失效分析和故障诊断、电动机和常用低压电器。

2. 现代设计方法及应用 创新设计、绿色产品设计、并行设计、虚拟设计、快速响应变型设计和反求设计、可靠性设计、摩擦学设计、优化设计、计算机辅助设计、疲劳强度设计、蠕变设计、价值工程、机电一体化系统设计、附录 现代设计主流软件。

3. 机械零部件设计 零部件设计常用基础标准、零件结构工艺性、联接与紧固、弹簧、起重、搬运件、操作件、机架、箱体及导轨、密封件、管路附件。

4. 机械传动设计 传动总论、机构、带传动和链传动、摩擦轮及螺旋传动、齿轮传动、轮系、减速器和变速器、轴、滚动轴承、滑动轴承、联轴器、离合器与制动器。

5. 流体传动与控制 气压传动与控制、液压传动与控制、液力传动。

为了便于协调,提高质量,加快编写进度,参加编审的人员以东北大学有关院系为主,并组织邀请清华大学、北京理工大学、北京科技大学、上海交通大学、上海大学、天津大学、哈尔滨工业大学、重庆大学、浙江大学、昆明理工大学、大连理工大学、大连铁道学院、华中理工大学、北京、上海、合肥、天津、沈阳等地的专家学者参加。值此手册出版之际,谨向所有参加本版工作的全体编审人员及有关单位表示诚挚的谢意。由于水平和时间有限,难免有一些不尽人意之处,殷切希望广大读者批评指正,提出宝贵意见,以便在今后的工作中改进。

# 目 录

## 第 36 篇 创新设计

### 第 1 章 创造与设计思维原理和方法

1 创新思维和方法	36-3
1.1 创新思维	36-3
1.2 创新思维机制	36-4
1.3 创新的外部因素与思维法则	36-4
1.4 创造活动的组织科学	36-6
2 基于功能的设计思维	36-8
2.1 功能的分类	36-8
2.2 功能分析	36-9
2.3 功能综合与整理	36-10
2.4 功能评价	36-11
3 基于产品寿命周期的设计思维	36-13
3.1 产品的市场寿命周期	36-13
3.2 产品的全寿命周期	36-14
4 基于过程的设计思维	36-14
4.1 设计任务的类型和进程	36-14
4.2 设计任务的形成与决策	36-16
4.3 方案设计	36-20
4.4 详细设计	36-23
5 产品总体设计的评价与优化	36-24
5.1 总体设计优化的目标与过程	36-24
5.2 优化对象的确定	36-24

5.3 总体设计的评价	36-28
5.4 总体设计的优化	36-33
6 方案设计和技术设计案例分析	36-35
6.1 家用单手把混水器方案设计案例分析	36-35
6.2 冲击试验台方案设计及详细设计案例分析	36-39

### 第 2 章 产品设计中的“人-机-环境”大系统观

1 “人-机-环境”大系统的构成与运行	36-58
1.1 “人-机-环境”大系统的构成	36-58
1.2 “人-机-环境”大系统的运行	36-58
2 功能分配	36-60
2.1 子系统能力的分析	36-60
2.2 功能分配原则	36-61
3 人的生理和心理因素	36-62
3.1 人的生理因素	36-62
3.2 人的心理因素	36-62
3.3 人的行为	36-64
4 环境因素与可持续发展观	36-65
参考文献	36-66

## 第 37 篇 绿色产品设计

### 第 1 章 概 述

1 绿色产品设计思想的由来	37-3
1.1 人类社会生存发展面临的三大威胁	37-3
1.1.1 人口问题	37-3
1.1.2 资源问题	37-3
1.1.3 环境污染问题	37-4
1.2 绿色产品设计的发展概况	37-6

1.2.1 国外的发展概况	37-6
1.2.2 国内的发展概况	37-6
1.3 实施绿色产品设计存在的问题	37-6
1.3.1 产品生产者存在的问题	37-6
1.3.2 产品的使用(消费)者存在的问题	37-8
1.4 节约资源和防止污染的策略	37-9
1.4.1 “源头预防”策略	37-9
1.4.2 “末端治理”策略	37-9

<b>第 2 章 绿色产品设计</b>	
2.1 概述 .....	37-10
2.2 传统设计与绿色设计 .....	37-10
2.2.1 传统产品设计 .....	37-10
2.2.2 绿色产品设计 .....	37-10
2.3 传统产品设计与绿色产品设计的 异同 .....	37-11
2.4 绿色产品设计的基本概念与 方法 .....	37-11
2.4.1 基本概念 .....	37-12
2.4.2 绿色产品设计的分析、评价 方法 .....	37-12
2.4.3 设计方法 .....	37-14
2.5 绿色产品设计的工具 .....	37-16
2.5.1 供设计人员评估其产品环境影 响用的定性准则 .....	37-16
2.5.2 生态指标法 .....	37-18
2.5.3 生命周期评估法(LCA) .....	37-22
<b>第 3 章 绿色产品设计中的设计准则</b>	
3.1 绿色产品设计中的结构设计准则 ...	37-26
3.2 绿色产品设计中的材料选择准则 ...	37-27
3.3 绿色产品设计中的工艺选择准则 ...	37-28
3.4 绿色产品设计中的产品使用阶段 的设计准则 .....	37-29
3.5 绿色产品设计中产品维修的设计 准则 .....	37-29
3.6 绿色产品设计中产品回收阶段的 设计准则 .....	37-30
<b>第 4 章 绿色产品设计的评价指标     系统及其评价方法</b>	
4.1 绿色设计评价指标体系 .....	37-32
4.2 绿色产品设计综合评价的方法 .....	37-32
4.2.1 层次分析法 .....	37-32
4.2.2 加权平均法 .....	37-35
4.2.3 模糊综合决策 .....	37-36
4.3 实例验证 .....	37-36
<b>第 5 章 绿色包装设计</b>	
5.1 产品包装设计的现状 .....	37-39
5.1.1 产品包装的基本功能 .....	37-39
5.1.2 产品包装引起的环境问题 .....	37-39
5.2 绿色包装设计 .....	37-40
5.2.1 绿色包装设计的内容和步骤 ...	37-40
5.2.2 绿色包装的设计准则 .....	37-41
<b>附 录</b>	
附录 1 废弃产品最终处置时受到限制的 材料 .....	37-43
附录 2 国外用于 LCA(生命周期分析) 的工具软件一览表 .....	37-43
附录 3 国外开发的用于产品生命周期分析 (LCA)的数据库(环境影响因素清 单分析用的数据) .....	37-44
附录 4 国外从事生态设计(Ecodesign)或 环境意识设计与制造(Environmentally Conscious Design and Manufacturing) 的机构名录 .....	37-45
附录 5 国外从事生命周期评估(LCA) 的机构名录 .....	37-45
附录 6 有关绿色产品设计的相关网站 .....	37-45
附录 7 塑料标记的标准 .....	37-50
参考文献 .....	37-51

## 第 38 篇 智能设计

### 第 1 章 智能模拟的科学

1 信息社会与思维科学 .....	38-3	1.2.2 形象(直觉)思维学 .....	38-5
1.1 思维与思维科学 .....	38-3	1.2.3 灵感(顿悟)思维学 .....	38-6
1.2 思维的类型 .....	38-3	2 思维的基础和认知的发展 .....	38-7
1.2.1 抽象(逻辑)思维学 .....	38-3	2.1 思维与智能 .....	38-7
		2.2 思维的神经基础 .....	38-7
		2.3 认知发展 .....	38-8

2.3.1	皮亚杰认知发展理论	38-8	1.1	遗传算法的概貌	38-28
2.3.2	斯腾伯格的认知三元素理论	38-9	1.2	单纯型遗传算法	38-29
2.3.3	信息加工理论	38-9	1.3	模式定理(schemata theorem)	38-31
2.3.4	思维的瞬间达尔文进化机制理论	38-9	1.4	遗传算法的有关操作规则和方 法	38-32
2.3.5	广义进化认知模式	38-9	1.5	多个体参与交叉的遗传算法	38-35
2.3.6	复杂自适应系统	38-10	1.6	多目标进化算法简介	38-39
2.3.7	认知发展总论	38-11	1.6.1	传统多目标算法及其存在 问题	38-39
3	智能模拟	38-11	1.6.2	Pareto 多目标进化算法	38-40
3.1	智能模拟的科学基础	38-11	1.6.3	几种主要的多目标进化算法	38-42
3.2	智能模拟的哲学基础	38-12	1.6.4	扩展 Pareto 进化算法(extended pareto evolutionary algorithm; EPEA)	38-44
3.3	智能模拟的基本途径	38-12	1.6.5	算例	38-46
3.3.1	基于逻辑推理的智能模拟—— 符号主义(Symbolism)	38-12	2	基于进化的健壮性设计方法	38-47
3.3.2	基于神经网络的智能模拟—— 联接主义(Connectionism)	38-12	2.1	健壮性开发方法的基本思路	38-47
3.3.3	基于“感知—行动”的智能模拟—— 行为主义(Behaviourism)	38-13	2.2	基于进化的健壮性设计方法的 总体框架	38-49
<b>第 2 章 智能设计方法和技术综述</b>			2.3	基于进化的健壮性设计方法的 说明	38-51
1	智能设计的发展概述	38-14	3	结构智能优化设计——进化设计	38-52
1.1	CAD 的发展	38-14	3.1	结构智能设计的概念	38-52
1.2	智能设计的两个阶段	38-14	3.2	结构进化智能优化设计	38-53
2	智能设计的概念和特征	38-15	3.3	基于进化的桁架结构相位设计	38-53
2.1	智能设计的特点	38-15	3.4	基于进化的结构非线性强制振动 解法	38-54
2.2	智能设计技术的研究重点	38-15	3.5	基于进化的圆抛物面天线健壮结 构设计	38-57
2.3	智能化方法的分类和智能设计的 层次	38-16	3.5.1	圆抛物面天线结构设计的要 求和特点	38-57
2.3.1	智能化方法的分类	38-16	3.5.2	天线反射面精度计算	38-58
2.3.2	智能设计的层次	38-16	3.5.3	最佳吻合抛物面各点对原设 计面相应点的半光程差	38-59
2.4	智能设计的基本方法	38-17	3.5.4	10m 圆抛物面天线健壮 设计模型	38-60
2.4.1	智能设计的分类	38-17	3.5.5	10m 圆抛物面天线体结构的 健壮性设计过程	38-61
2.4.2	智能设计系统与技术	38-18	3.5.6	总结	38-69
3	智能设计体系和知识表达	38-19	4	供应链库存策略的进化重组	38-70
3.1	智能设计体系	38-19	4.1	供应链运行策略的持续改进	38-70
3.1.1	智能设计的抽象层次模型	38-19	4.2	供应链中的库存设置	38-71
3.1.2	设计知识的结构体系	38-20	4.3	供应链运行过程中的库存控制 策略	38-72
3.1.3	智能设计的集成求解策略	38-21	4.4	敏捷供应链多级库存策略重组 模型	38-74
3.1.4	智能设计集成求解策略工程 应用	38-22			
3.2	智能设计的知识表达	38-23			
3.3	智能设计的基因模型表达	38-26			
3.3.1	知识模型	38-26			
3.3.2	基因模型	38-26			
<b>第 3 章 进化设计技术与方法</b>					
1	进化设计技术基础	38-28			

## 第 4 章 自组织设计技术与方法

- 1 自组织技术基础 ..... 38-80
  - 1.1 “生命的游戏” ..... 38-80
  - 1.2 元胞自动机的基础 ..... 38-81
  - 1.3 元胞自动机的自组织建模方法 ..... 38-84
  - 1.4 元胞自动机的应用领域 ..... 38-85
- 2 结构拓扑的自组织进化 ..... 38-86
  - 2.1 结构拓扑优化中的 ECA 直接规则 ..... 38-87
  - 2.2 ECA 规则的进化表达 ..... 38-89
  - 2.3 结构拓扑形态优化的算例 ..... 38-89

## 第 5 章 自学习设计技术与方法

- 1 自学习技术基础 ..... 38-91
  - 1.1 神经网络的概述 ..... 38-91
  - 1.2 神经网络的主要特点 ..... 38-92
  - 1.3 细胞元模型 ..... 38-93
  - 1.4 神经网络模型 ..... 38-95
  - 1.5 神经网络的学习 ..... 38-96
  - 1.6 多层前向神经网络(BP 网络) ..... 38-99
  - 1.7 典型反馈网络——Hopfield 网络 ..... 38-105
  - 1.8 基于概率学习的 Boltzmann 机模型 ..... 38-107
- 2 非线性振动的自学习建模 ..... 38-110
  - 2.1 神经网络和系统识别 ..... 38-110
  - 2.2 非线性振动脉冲响应的学习和系统预测 ..... 38-112
  - 2.3 Duffing 振动的学习和预测 ..... 38-112
  - 2.4 预测精度和泛用性的考察 ..... 38-115
- 3 基于学习的机械系统特性预测 ..... 38-117
  - 3.1 机械系统特性预测的问题 ..... 38-117
  - 3.2 机械系统特性预测的基本模型 ..... 38-117
  - 3.3 雷达结构系统固频的预测例 ..... 38-118
- 4 神经网络专家系统的智能设计体系结构 ..... 38-119

- 1.1 建立人工神经网络专家系统的必要性 ..... 38-119
- 4.2 面向设计的智能平台 ..... 38-119
  - 4.2.1 专家系统和神经网络的结合方式 ..... 38-119
  - 4.2.2 智能平台的“外壳”结构 ..... 38-119
  - 4.2.3 设计求解过程 ..... 38-120
  - 4.2.4 知识的处理方法 ..... 38-120
- 4.3 说明 ..... 38-120
- 5 基于神经网络的 CAD/CAM 一体化 ..... 38-121
  - 5.1 系统的结构 ..... 38-121
  - 5.2 产品零件数据结构 ..... 38-121
  - 5.3 智能 CAPP 系统 ..... 38-121
    - 5.3.1 BP 网络实现加工链的选择 ..... 38-121
    - 5.3.2 工艺尺寸链计算的 Hopfield 网络 ..... 38-122
  - 5.4 CAM 模块 ..... 38-123

## 第 6 章 人工生命设计技术与方法

- 1 人工生命技术基础 ..... 38-124
  - 1.1 人工生命的进化模型 ..... 38-124
  - 1.2 L 系统与形态生成模型 ..... 38-127
- 2 人工生命的研究内容归纳 ..... 38-128
  - 2.1 数字生命的研究 ..... 38-128
  - 2.2 数字社会的研究 ..... 38-129
  - 2.3 虚拟生态环境 ..... 38-129
  - 2.4 人工脑(Artificial Brain) ..... 38-129
  - 2.5 进化机器人(evolutionary robotics) ..... 38-129
  - 2.6 进化软件代理(evolvable multiagent) ..... 38-130
- 3 人工生命的设计方法 ..... 38-130
  - 3.1 金融证券市场分析决策中的人工生命应用 ..... 38-130
  - 3.2 计算机动画的人工生命应用 ..... 38-131
  - 3.3 基于人工生命的因特网提速 ..... 38-132
- 参考文献 ..... 38-133

## 第 39 篇 并行设计

### 第 1 章 并行设计概述

- 1 并行工程概念与方法学 ..... 39-3
  - 1.1 并行工程概念 ..... 39-3
  - 1.2 并行工程方法学 ..... 39-5
- 2 并行设计特点与本质 ..... 39-5
  - 2.1 并行设计特点 ..... 39-5
  - 2.2 并行设计本质 ..... 39-6

3 并行设计中的关键技术 .....	39-6
--------------------	------

## 第 2 章 并行产品开发过程 建模及冲突的预消解

1 并行产品开发过程模型 .....	39-8
1.1 并行产品开发过程的定义 .....	39-8
1.2 并行设计过程的基本元素 .....	39-8
1.2.1 活动(Activity) .....	39-8
1.2.2 成员(Person) .....	39-9
1.2.3 角色(Role) .....	39-9
1.2.4 资源(Resource) .....	39-10
1.2.5 产品数据(Product Data) .....	39-10
1.3 并行设计过程的基本视图 .....	39-11
1.3.1 并行设计过程基本元素之间的 关系——子视图 .....	39-11
1.3.2 活动流视图 .....	39-12
1.3.3 组织视图 .....	39-12
1.3.4 资源视图 .....	39-15
1.3.5 产品数据视图 .....	39-15
1.4 并行设计过程的递阶集成 多视图模型 .....	39-16
2 并行设计过程中冲突的预消解 .....	39-17
2.1 冲突分析 .....	39-17
2.1.1 冲突的形成 .....	39-17
2.1.2 冲突的概念与内涵 .....	39-17
2.1.3 冲突产生的原因 .....	39-18
2.1.4 冲突的分类 .....	39-19
2.2 冲突的预消解 .....	39-19

## 第 3 章 workflow 管理系统

1  workflow 的基本概念 .....	39-22
1.1  workflow 的定义 .....	39-22
1.2  workflow 基本术语 .....	39-22
1.3  workflow 管理系统类型 .....	39-24
2  workflow 管理系统参考模型 .....	39-25
2.1 过程定义 .....	39-25
2.1.1 过程定义工具 .....	39-25
2.1.2  workflow 定义转换(接口 1) .....	39-25
2.1.3 基本元模型(A Basic Meta- Model) .....	39-26
2.2  workflow 客户端功能 .....	39-26
2.2.1  workflow 客户端应用程序(Workflow Client Applications) .....	39-26
2.2.2  workflow 客户端应用程序接口 (接口 2) .....	39-26

2.3 应用程序调用功能 .....	39-27
2.3.1 应用程序调用 .....	39-27
2.3.2 应用程序调用接口 (接口 3) .....	39-27
2.4  workflow 互操作能力 .....	39-27
2.4.1 协同工作模式 .....	39-27
2.4.2  workflow 互操作分类 .....	39-29
2.4.3  workflow 互操作的实现技术 .....	39-29
2.4.4  workflow 管理系统互操作性 存在的问题 .....	39-30
2.5 系统管理 .....	39-31
2.6  workflow 故障恢复和异常处理 .....	39-31
2.7  workflow 规范 .....	39-31
3  workflow 管理技术研究现状 .....	39-32
3.1  workflow 建模技术的研究现状 .....	39-32
3.2  workflow 管理系统的研究现状 .....	39-32
3.3  workflow 技术的不足 .....	39-34
4  workflow 产品简介 .....	39-34
4.1 FileNet 公司的工作流产品 .....	39-34
4.2 JetForm 公司的工作流产品 .....	39-35
4.3 IBM 的工作流产品 .....	39-35
4.4 Action 技术公司的工作流产品 .....	39-35
5  workflow 管理系统的发展趋势 .....	39-36

## 第 4 章 产品数据交换技术

1 产品数据标准概述 .....	39-37
1.1 产品数据及产品模型 .....	39-37
1.2 产品数据交换接口 .....	39-37
1.3 产品数据标准的发展 .....	39-38
2 产品模型数据标准 STEP .....	39-40
2.1 STEP 标准的特点 .....	39-40
2.2 STEP 的体系结构 .....	39-40
2.3 EXPRESS 描述语言 .....	39-42
2.4 STEP 标准化状况 .....	39-43
3 主要应用协议介绍 .....	39-43
3.1 电器设计和装配应用协议 AP212 .....	39-43
3.2 汽车机械设计过程 AP214 .....	39-44
3.3 工艺设计的机加工产品定义应用 协议 AP224 .....	39-44
4 STEP 工业应用 .....	39-45
4.1 Pro STEP 中心 .....	39-45
4.2 AP212 和 AP214 的工业推广 .....	39-46
4.2.1 AP212 工业推广 .....	39-46
4.2.2 AP214 工业推广 .....	39-46

<b>第 5 章 并行设计的使能工具</b>	
1 质量功能配置(QFD, Quality Function Deployment) .....	39-47
2 面向制造的设计(DFM, Design for Manufacturing) .....	39-48
2.1 基于特征的零件信息模型 .....	39-48
2.2 特征信息提取和匹配 .....	39-49
2.3 产品可制造性评价 .....	39-50
3 面向装配的设计(DFA, Design for Assembly) .....	39-51
3.1 产品结构树 .....	39-51
3.2 装配模型 .....	39-52
3.3 可装配性评价 .....	39-52
4 DFM 技术在摆线针轮减速器设计上的应用 .....	39-52
4.1 摆线针轮减速器设计应用 DFM 技术系统结构 .....	39-52
4.2 产品总体设计 .....	39-53
4.3 零件特征造型 .....	39-53
4.4 装配分析 .....	39-53
4.5 零件 DFM 评价 .....	39-53
<b>第 6 章 并行设计的集成平台 PDM</b>	
1 PDM 概述 .....	39-55
1.1 PDM 的产生背景 .....	39-55
1.2 PDM 的概念 .....	39-55
1.3 PDM 的特点 .....	39-56
1.4 PDM 的一般体系结构 .....	39-56
2 PDM 的主要功能 .....	39-57
3 PDM 的集成 .....	39-58
3.1 PDM 集成的种类 .....	39-58
3.2 PDM 与应用系统集成的层次 .....	39-58
3.3 PDM 对数据信息集成的方法 .....	39-60
3.4 基于接口模式的应用集成策略 .....	39-60
4 PDM 的实施 .....	39-61
<b>第 7 章 并行设计实例</b>	
1 基于 PDM 的油泵产品并行设计原型系统 .....	39-62
1.1 系统目标 .....	39-62
1.2 体系结构 .....	39-62
1.3 IPDT 设计 .....	39-62
1.4 电子仓库设计 .....	39-63
1.5 产品开发过程管理 .....	39-63
1.7 基于 PDM 的油泵产品并行开发原型系统 .....	39-64
1.7.1 油泵产品并行开发过程的活动流 .....	39-64
1.7.2 产品开发过程监控器功能 .....	39-66
2 基于 PDM 的客车车身并行设计 .....	39-69
2.1 客车车身并行设计开发项目概况 .....	39-69
2.2 PDM 系统结构 .....	39-70
2.3 逻辑拓扑 .....	39-70
2.4 文件系统定义 .....	39-70
2.5 Client/Server 与 Web/Browser 结构 .....	39-70
2.6 用户、角色定义 .....	39-71
2.7 Vault 结构及权限控制 .....	39-72
2.8 流程管理 .....	39-75
2.9 多反馈机制及实现 .....	39-83
2.10 工具封装 .....	39-86
参考文献 .....	39-87

## 第 40 篇 有限元法及其工程应用

### 第 1 章 弹性力学问题有限元法原理和表达式

1 广义坐标有限元法的一般格式 .....	40-3
1.1 选择单元位移函数的一般原则 .....	40-3
1.2 广义坐标有限元的一般格式 .....	40-3
2 平面问题三角形单元的有限元格式 .....	40-3

3 平面问题矩形单元的有限元格式 .....	40-5
4 轴对称问题的有限元格式 .....	40-7
5 空间问题的有限元格式 .....	40-8

### 第 2 章 等参单元与数值积分

1 一维等参单元 .....	40-10
2 二维等参单元 .....	40-10
3 三维等参单元 .....	40-11

4 等参数单元用十弹性力学分析的一般格式	40-14
5 数值积分方法	40-14
5.1 一维数值积分	40-14
5.2 二维及三维高斯积分	40-15
6 等参单元计算中数值积分阶次的选择	40-15

### 第3章 杆件结构力学问题的有限单元法

1 等截面直杆—梁单元	40-16
1.1 轴力杆单元	40-16
1.2 扭转杆单元	40-16
1.3 弯曲梁单元	40-16
2 杆件系统结构分析	40-17
2.1 杆系结构的矩阵分析	40-17
2.2 平面杆系结构的分析公式	40-17
2.3 空间杆系结构	40-17

### 第4章 平板弯曲问题的有限单元法

1 非协调板单元	40-19
1.1 矩形板单元	40-19
1.2 三角形板单元	40-20
2 协调板单元	40-20
3 Mindlin 板单元	40-21
4 应力杂交板单元	40-22

### 第5章 轴对称壳体问题的有限单元法

1 基于薄壳理论的轴对称壳体单元	40-23
1.1 轴对称薄壳理论的基本公式	40-23
1.2 薄壳截锥单元	40-24
2 位移和转动各自独立插值的轴对称壳体单元	40-25
2.1 考虑横向剪切变形的轴对称壳体理论的基本公式	40-25
2.2 截锥单元	40-25
2.3 曲边单元	40-26
3 轴对称超参数壳体单元	40-26
3.1 几何形状的规定	40-26
3.2 位移函数	40-27
3.3 应变和应力的确定	40-27
3.4 刚度矩阵和节点载荷的计算	40-27
4 不同类型的单元的联结	40-28
4.1 多点约束方程	40-28
4.1.1 罚函数法	40-28
4.1.2 直接引入法	40-28

4.2 过渡单元	40-28
----------	-------

### 第6章 一般壳体问题的有限单元法

1 平板壳体单元	40-29
1.1 局部坐标系内的单元刚度矩阵	40-29
1.2 单元刚度矩阵从局部坐标系列整体坐标系的转换	40-29
2 超参数壳体单元	40-30
2.1 几何形状的规定	40-30
2.2 位移函数的表示	40-30
2.3 应变和应力的确定	40-30
2.4 刚度矩阵和节点载荷的计算	40-31
3 相对自由度壳体单元	40-31
4 不同类型单元的联结	40-32
4.1 多点约束方程	40-32
4.2 过渡单元	40-32

### 第7章 热传导问题的有限单元法

1 稳态热传导问题	40-33
2 瞬态热传导问题	40-34
3 热应力的计算	40-35

### 第8章 动力学问题的有限单元法

1 质量矩阵与阻尼矩阵	40-36
1.1 协调质量矩阵和集中质量矩阵	40-36
1.2 振型阻尼矩阵	40-36
2 直接积分法	40-36
2.1 中心差分法	40-37
2.2 Newmark 方法	40-37
3 振型叠加法	40-37
3.1 将运动方程转换到正则振型坐标系	40-37
3.2 求解单自由度系统振动方程	40-38
3.3 振型叠加得到系统的响应	40-38
4 大型特征值问题的解法	40-38
4.1 反迭代法	40-38
4.2 子空间迭代法	40-38
4.3 Ritz 向量直接叠加法	40-38
5 减缩系统自由度的方法	40-38
5.1 主从自由度法	40-38
5.2 模态综合法	40-38

### 第9章 材料非线性问题的有限单元法

1 材料非线性本构关系	40-39
2 弹塑性增量分析有限元格式	40-39



3	非线性方程组的解法	40-40
3.1	直接迭代法	40-40
3.2	牛顿迭代法	40-40
3.3	修正牛顿法	40-40

### 第 10 章 几何非线性问题的有限单元法

1	大变形情况下的应变和应力	40-41
1.1	应变的度量	40-41
1.2	应力的度量	40-42
2	几何非线性问题的表达格式	40-43
3	大变形条件下的本构关系	40-43
3.1	弹性	40-43
3.1.1	大位移、大转动、小应变情况	40-43
3.1.2	大应变情况	40-43
3.2	非弹性	40-44
3.2.1	大位移、大转动、小应变情况	40-44
3.2.2	大变形(包含大位移、大转动)情况	40-44
4	几何非线性问题的求解方法	40-44

### 第 11 章 有限单元法算例

1	空间桁架问题	40-45
---	--------	-------

2	平板弯曲问题	40-46
3	轴对称壳体问题	40-47
4	一般壳体问题	40-47
5	热传导问题	40-47
6	动力学问题	40-48
7	材料非线性问题	40-49
8	几何非线性问题	40-50

### 第 12 章 典型有限元软件简介

1	MSC. NASTRAN	40-51
2	ANSYS	40-51
3	LS-DYNA	40-51
4	国产有限元软件	40-51
4.1	GHFEA	40-52
4.2	有限元分析与结构优化微机软件系统 JFX95-W	40-52
4.3	有限元分析和优化设计系统 MAS	40-52
4.4	线性与非线性结构分析程序/交互式有限元图形分析系统 APOLANS/INFE GAS	40-52
	参考文献	40-52

## 第 41 篇 虚拟设计

### 第 1 章 虚拟设计总论

1	虚拟设计及其相关概念	41-3
1.1	虚拟制造	41-3
1.2	虚拟产品	41-3
1.3	虚拟设计	41-3
1.4	虚拟设计与其他概念之间的关系	41-4
2	虚拟设计结构体系	41-5
3	数字化辅助技术的发展趋势	41-6

### 第 2 章 虚拟现实技术

1	虚拟现实的特点	41-7
2	虚拟现实系统组成及分类	41-7
2.1	虚拟现实系统的组成	41-7
2.2	虚拟现实系统的分类	41-7
3	产生虚拟现实环境的工具集	41-8
3.1	产生虚拟现实环境的硬件	41-8
3.1.1	虚拟现实系统的硬件组成	41-9

3.1.2	可获得的硬件系统	41-11
3.2	虚拟现实的软件子系统	41-12
3.2.1	虚拟现实软件系统的组成	41-12
3.2.2	可获得的虚拟现实软件系统	41-13

### 第 3 章 基于虚拟现实技术的新一代 CAD 技术

1	基于虚拟现实的 CAD 的特点	41-15
2	VR-CAD 的几何建模技术	41-15
2.1	传统的几何建模技术简介	41-15
2.2	VR-CAD 中的几何建模	41-16
3	VR-CAD 中的多通道技术	41-18
3.1	三维鼠标	41-18
3.2	三维物体选取机制	41-18
3.3	三维菜单的设计	41-19
3.4	语音系统	41-19
3.5	触觉和力觉反馈系统	41-20

4	VR—CAD 中的可视化技术	41—21
4.1	VR—CAD 中真实感图形实时绘制技术	41—21
4.2	VR—CAD 中多细节程度模型生成技术	41—22
4.2.1	网格简化	41—22
4.2.2	多分辨率模型生成	41—23
4.3	VR—CAD 系统中的复杂场景实时漫游技术	41—23
5	设计中的人员因素分析	41—24

#### 第 4 章 虚拟现实开发工具集

1	虚拟现实建模语言(VRML)	41—27
1.1	VRML 的概述和发展历程	41—27
1.2	VRML2.0 的语言规范	41—28
1.2.1	VRML 文件	41—28
1.2.2	VRML 的节点及用法	41—30
1.2.3	VRML 的细节层次控制	41—33
1.2.4	VRML 的纹理映射及控制	41—34
1.2.5	VRML 的光照和雾化	41—34
1.2.6	VRML 的背景及添加声音	41—35
1.2.7	VRML 的视点控制	41—36
1.2.8	VRML 中程序脚本的灵活应用	41—36
1.3	VRML 的应用及常用资源	41—36
1.3.1	典型的应用领域	41—36
1.3.2	成功案例	41—37
1.3.3	VRML 浏览器	41—37
1.3.4	VRML 常用资源	41—38
2	实时视景开发工具——OpenGL Performer 基础	41—38
2.1	OpenGL Performer 基础	41—38
2.1.1	组成	41—38
2.1.2	渲染流程	41—39
2.1.3	节点和节点类型	41—40
2.1.4	数据载入	41—41
2.1.5	可视化数据库的遍历	41—42
2.1.6	帧和载入控制	41—42
2.1.7	可视化效果	41—43
2.2	基于 Performer 的实时仿真系统开发框架	41—44
2.2.1	场景创建的基本框架	41—44
2.2.2	基于 libpf 应用程序的结构分析	41—45
2.2.3	基于 libpfv 应用程序的结构分析	41—47

#### 第 5 章 工程分析数据沉浸可视化技术

1	工程分析数据沉浸可视化技术概述	41—52
1.1	科学计算可视化的概念	41—52
1.2	科学计算可视化的研究内容	41—52
1.3	科学计算可视化技术发展趋势	41—52
2	数据模型准备	41—53
2.1	CAD 数据的精简	41—53
2.2	CAE 数据的转换	41—55
3	科学计算可视化的基础技术	41—56
3.1	可视化数据的组织形式及物理分类	41—56
3.2	可视化流程	41—57
3.3	矢量场特征可视化	41—58
3.3.1	基于拓扑结构的矢量场特征可视化	41—59
3.3.2	种子点选取和布置方法	41—60
3.3.3	矢量场特征可视化实例	41—60
4	矢量场数据的沉浸可视化关键技术	41—61
4.1	适合各种网格单元的自适应步长快速流线生成算法	41—61
4.2	流面生成算法	41—64
4.3	流管生成算法	41—65
4.4	三维实体图标	41—65
4.5	颜色映射方法	41—66

#### 第 6 章 虚拟装配技术

1	虚拟装配的概述及其国内外研究简介	41—67
1.1	概述	41—67
1.2	国内外研究简介	41—67
2	虚拟装配关键技术	41—69
2.1	虚拟装配模型	41—69
2.1.1	传统的装配模型	41—69
2.1.2	虚拟环境下的装配建模	41—69
2.2	虚拟装配过程	41—70
2.3	碰撞检测	41—71
2.3.1	虚拟装配中的碰撞检测	41—71
2.3.2	公开算法软件包	41—71
2.4	虚拟装配路径规划和仿真	41—72
3	典型虚拟装配系统功能介绍	41—72

#### 第 7 章 面向产品开发的工程应用

1	应用背景介绍	41—74
2	虚拟客车车身开发系统的集成平台	41—74

2.1 数据仓库规划 .....	41-75	分析 .....	41-81
2.2 流程管理 .....	41-75	4.4 大客车整车的碰撞安全性数值	
3 车身外形设计与性能分析 .....	41-77	模拟 .....	41-81
3.1 客车车身曲面设计 .....	41-77	5 内饰设计与性能分析 .....	41-82
3.2 客车车身外形性能分析 .....	41-78	5.1 司机座椅分析模型的建立 .....	41-82
4 结构设计与性能分析 .....	41-79	5.2 座椅的动态舒适性分析 .....	41-82
4.1 基于三维模型的车身结构设计及		6 虚拟客车样机和人机工程分析 .....	41-83
装配 .....	41-79	6.1 虚拟客车样机的建立 .....	41-83
4.2 基于三维模型的车身结构静强度		6.2 人机工程分析 .....	41-84
分析 .....	41-80	参考文献 .....	41-85
4.3 基于三维模型的车身结构模态			

## 第 42 篇 快速响应变型设计

### 第 1 章 绪 论

1 快速响应工程及其含义 .....	42-3
2 快速响应设计是实施快速响应	
工程的重要一环 .....	42-4
3 用变型设计实现快速响应的方法 .....	42-4

### 第 2 章 快速响应变型设计的体系结构

1 变型设计的过程 .....	42-7
2 变型设计的层次结构 .....	42-8
3 变型设计的系统框架 .....	42-9

### 第 3 章 快速响应变型设计的关键技术

1 事物特性表管理 .....	42-12
1.1 事物特性技术简介 .....	42-12
1.2 事物特性 .....	42-12
1.3 特性描述的对象 .....	42-14
1.4 事物特性表(Layouts of Article	
Characteristics) .....	42-14
1.5 事物特性技术及其特点 .....	42-15
2 分类编码系统 .....	42-15
2.1 零件分类编码系统的基本原理	
和结构 .....	42-15
2.2 适用于快速响应设计的分类编	
码系统 .....	42-16
3 产品资源管理 .....	42-17
3.1 产品资源管理的意义 .....	42-17
3.2 产品资源管理中的设计信息模型 .....	42-18
3.3 产品信息管理 .....	42-20
3.4 产品后台信息的定义与管理 .....	42-21

4 关系型 CAD 系统 .....	42-23
4.1 传统 CAD 系统的局限性 .....	42-23
4.2 参数化设计和变量化设计 .....	42-24
4.3 适用于变型设计的关系型 CAD	
系统 .....	42-25

### 第 4 章 关系型产品模型理论及其应用

1 关系型产品模型的定义及应用背景 .....	42-29
1.1 变型设计在订单规划中的作用 .....	42-29
1.2 企业产品信息资源重组 .....	42-30
2 对象类的信息构成 .....	42-32
2.1 产品对象定义 .....	42-32
2.2 各产品对象类的信息构成 .....	42-36
3 对象分类框架 .....	42-36
3.1 GT 分类原理 .....	42-37
3.2 对象分类框架 .....	42-38
4 基于关系的族类属模型 .....	42-41
4.1 对象类的基本特性 .....	42-41
4.2 对象类的类属模型 .....	42-42

### 第 5 章 基于实例推理的快速响应变型设计

1 CBR 概述 .....	42-44
2 CBR 的基本问题 .....	42-47
2.1 实例库的建立 .....	42-47
2.2 实例的检索和提取方法 .....	42-48
2.3 实例的修改 .....	42-48
3 应用于快速响应变型设计的 CBR	
关键技术研究 .....	42-49
3.1 基本概念 .....	42-49
3.2 实例、实例原型的关联关系 .....	42-50

3.3 实例、实例原型的表示、索引和组织原理.....	42-50	4 基于 CBR 的产品快速响应变型设计 .....	42-59
3.3.1 实例和实例原型表示.....	42-50	4.1 产品定义变型设计.....	42-60
3.3.2 实例和实例原型的索引、组织 .....	42-52	4.2 产品装配变型设计.....	42-60
3.3.3 实例的检索.....	42-53	4.3 产品概念变型设计.....	42-61
3.4 实例的修改.....	42-57	<b>第 6 章 实例研究</b>	
3.4.1 实例修改的基本问题和修改策略.....	42-58	1 关系型产品模型的建模步骤.....	42-62
3.4.2 基于约束满足技术的实例修改过程.....	42-59	2 圆锥—圆柱行星齿轮减速机的变型设计.....	42-62
		3 基于产品族的液压泵变型设计实例.....	42-63

## 第 43 篇 模块化设计

### 第 1 章 概 述

1 现代设计的特征 .....	43-3
1.1 设计的概念 .....	43-3
1.2 设计思想的演变 .....	43-4
1.3 现代设计的特征 .....	43-5
2 模块化设计的由来 .....	43-5

### 第 2 章 模块化设计的基础——功能分析

1 功能分析的概念 .....	43-7
2 功能的分类 .....	43-8
3 功能分析的涵义 .....	43-8
4 功能结构图 .....	43-10
4.1 功能结构 .....	43-10
4.2 功能结构图的建立 .....	43-10

### 第 3 章 模块化设计基础

1 模块与模块化 .....	43-12
1.1 什么是模块化(Modularization) .....	43-12
1.2 模块化的作用 .....	43-13
2 模块的划分原则 .....	43-13
3 模块的组合原则 .....	43-15
4 模块的接口设计 .....	43-15
5 模块化与标准化 .....	43-16

### 第 4 章 模块化产品的设计过程

1 设计过程 .....	43-17
2 需求分析 .....	43-17
3 可重构设计的概念和模型 .....	43-18
4 设计中的单元重用 .....	43-18

1.1 重构单元的概念和分类 .....	43-18
4.2 单元重构过程 .....	43-19
4.2.1 RPD 概念 .....	43-19
4.2.2 RPD 流程 .....	43-20
4.2.3 RPD 关键技术 .....	43-20
4.3 优化的零件单元选择技术 .....	43-21
4.3.1 零件选择策略 .....	43-21
4.3.2 最强约束策略和最小损失策略 .....	43-21
4.3.3 零件选择算法 .....	43-22
4.4 应用举例 .....	43-22
5 量化法在设计过程规划中的应用 .....	43-24
5.1 量化法与设计过程规划 .....	43-25
5.1.1 设计评审规划技术及其研究 .....	43-25
5.1.2 量化法及相关研究 .....	43-25
5.2 基于量化法的任务影响因子模型 .....	43-25
5.2.1 模型原理 .....	43-25
5.2.2 模型计算步骤 .....	43-25
5.2.3 几点说明 .....	43-28
5.3 基于任务影响因子模型的设计评审规划 .....	43-28
5.3.1 需求量化的基本思想 .....	43-28
5.3.2 步骤 .....	43-28
5.3.3 应用举例 .....	43-28

### 第 5 章 模块化技术的应用

1 概述 .....	43-31
2 模块化在造船业中的应用 .....	43-32
3 模块化在临床医学中应用 .....	43-33
参考文献 .....	43-34

## 第 44 篇 优化设计

### 第 1 章 优化设计算法原理

1 优化设计概念 .....	44-3
1.1 优化设计基本概念 .....	44-3
1.2 优化数学模型 .....	44-3
1.3 优化算法分类 .....	44-4
2 一维无约束优化方法 .....	44-5
2.1 切线法(牛顿法) .....	44-5
2.2 黄金分割法(0.618 法) .....	44-6
2.3 插值法 .....	44-6
3 多维无约束优化算法 .....	44-7
3.1 坐标轮换法 .....	44-7
3.2 最速下降法(Cauchy 法, $n$ -阶梯 度法) .....	44-7
3.3 牛顿法 .....	44-7
3.4 变尺度法 .....	44-8
4 简约梯度法及广义简约梯度法 .....	44-8
4.1 简约梯度法 .....	44-8
4.2 广义简约梯度法 .....	44-9
5 罚函数法 .....	44-11
5.1 内点法 .....	44-11
5.2 外点法 .....	44-12
5.3 混合法 .....	44-13
5.4 增广拉格朗日乘子法 .....	44-14
6 序列线性规划法 .....	44-14
7 序列二次规划法 .....	44-15
8 遗传算法 .....	44-16
9 神经网络优化方法 .....	44-18
9.1 神经网络优化方法基本思路 .....	44-18
9.2 神经网络优化方法 .....	44-18

### 第 2 章 机构优化设计

1 机构优化设计概述 .....	44-21
2 连杆机构优化设计 .....	44-22
3 凸轮机构优化设计 .....	44-24
4 机构多目标优化设计 .....	44-25

### 第 3 章 机械零件优化设计

1 机械零件优化设计概述 .....	44-29
2 齿轮传动优化设计 .....	44-31
3 弹簧优化设计 .....	44-33

4 轴承优化设计 .....	44-35
----------------	-------

### 第 4 章 机械系统优化设计

1 机械系统优化设计概述 .....	44-37
2 机械系统动态优化设计 .....	44-37
3 机电系统优化设计 .....	44-39

### 第 5 章 结构优化设计

1 结构优化设计概述 .....	44-43
2 结构优化设计的准则法 .....	44-44
2.1 满应力法 .....	44-44
2.2 单位移约束准则法 .....	44-44
2.3 多位移约束准则法 .....	44-45
3 结构优化的齿行法 .....	44-45
3.1 结构优化齿行法简介 .....	44-45
3.2 杆结构优化齿行法 .....	44-46
3.3 梁结构优化齿行法 .....	44-46
1 结构优化的敏度分析技术 .....	44-47
4.1 敏度分析方法 .....	44-47
4.2 敏度分析的实现 .....	44-48
5 实例 .....	44-50

### 第 6 章 形状优化设计

1 形状优化设计概述 .....	44-51
2 形状优化的敏度分析 .....	44-52
2.1 位移敏度计算 .....	44-52
2.2 应力敏度计算 .....	44-53
2.3 形状敏度分析实现 .....	44-53
3 形状优化的自适应分析技术 .....	44-54
3.1 误差估计 .....	44-54
3.2 $h$ 自适应法 .....	44-55
3.3 $p$ 自适应法 .....	44-55
4 自适应分析形状优化设计 .....	44-57
5 形状优化设计实例 .....	44-58
5.1 连杆的形状优化设计 .....	44-58
5.2 起重吊钩的自适应形状优化 .....	44-59

### 第 7 章 可靠性优化设计

1 可靠性优化设计概述 .....	44-60
2 机构可靠性优化设计 .....	44-61

3 结构可靠性优化设计	44-63
-------------	-------

## 第 8 章 复杂系统优化设计

1 复杂系统优化设计概述	44-66
2 若干优化新算法	44-67
2.1 粒子群优化算法(PSO)	44-67

2.2 智能体(Agent)	44-67
2.3 免疫算法	44-69
2.4 蚁群算法	44-69
3 多级优化设计技术	44-70
4 多学科优化设计	44-71

# 第 45 篇 可靠性设计

## 第 1 章 产品可靠性及可靠性设计

1 产品质量与可靠性	45-3
1.1 产品质量	45-3
1.2 产品可靠性与有效性	45-3
1.3 产品全寿命周期费用	45-3
1.4 可靠性设计	45-4
2 可靠性问题的统计描述和表达	45-4
2.1 直方图和密度函数	45-4
2.2 分布函数和失效概率	45-5
2.3 可靠度	45-6
2.4 失效率	45-6
3 统计学参数	45-8
4 可靠性问题示例	45-9
5 产品可靠性指标	45-10
5.1 平均寿命	45-10
5.2 寿命方差与标准差	45-10
5.3 可靠寿命、中位寿命与特征寿命	45-10
5.4 维修度	45-11
5.5 有效度	45-11
5.6 可靠性特征量间的关系	45-11
6 可靠性设计程序和内容	45-12

## 第 2 章 可靠性设计的数学工具

1 可靠性设计常用的概率分布	45-14
1.1 二项分布	45-14
1.2 泊松(Poisson)分布	45-14
1.3 正态(Gauss)分布	45-14
1.4 对数正态分布	45-14
1.5 威布尔(Weibull)分布	45-15
1.6 指数分布	45-15
1.7 1型极值分布(Gumbel 分布)	45-15
1.8 $\Gamma$ 分布、瑞利分布、 $\beta$ 分布	45-16
1.9 $\chi^2$ 分布、 $t$ 分布、 $F$ 分布	45-16
2 假设检验	45-29

2.1 分布类型的假设检验	45-30
2.2 $\chi^2$ 检验法	45-30
2.3 K-S检验法	45-31
3 分布参数的估计	45-32
3.1 点估计	45-32
3.2 区间估计	45-32
3.3 极大似然估计	45-33

## 第 3 章 机械零件可靠性设计

1 机械零件可靠性设计的特点、内容与 方法	45-35
1.1 机械零件可靠性设计的特点	45-35
1.2 机械零件可靠性设计的基本内容及 步骤	45-35
2 应力-强度干涉模型与可靠度计算方法	45-35
2.1 应力-强度干涉模型	45-35
2.2 应力分布的确定	45-35
2.2.1 零件尺寸误差	45-35
2.2.2 应力分布参数的代数运算	45-36
2.3 强度分布	45-36
2.4 可靠度计算方法	45-37
2.4.1 应力-强度干涉模型求可靠度的 一般公式	45-37
2.4.2 数值积分法求可靠度	45-38
2.4.3 梅林变换法求可靠度	45-38
2.4.4 蒙特卡罗法	45-40
2.4.5 有多种失效模式的可靠度	45-43
3 可靠度的置信度和置信区间	45-43
3.1 置信度	45-43
3.2 单侧置信区间下限和最低可靠度	45-43
4 随机变量函数的均值和标准差的近似 计算	45-46
4.1 泰勒展开法	45-46
4.2 变异系数法	45-47

4.3 基本函数法 .....	45-47	5.1 基本概念 .....	45-74
5 典型机械零件的可靠性设计 .....	45-48	5.2 基本符号 .....	45-74
5.1 螺纹联接的可靠性设计 .....	45-48	5.2.1 事件符号 .....	45-74
5.1.1 松螺栓联接 .....	45-48	5.2.2 逻辑门符号 .....	45-75
5.1.2 紧螺栓联接 .....	45-49	5.3 故障树的建立 .....	45-75
5.1.3 受剪螺栓联接 .....	45-51	5.3.1 注意事项 .....	45-76
5.1.4 按栓杆或孔壁受挤压进行 设计 .....	45-52	5.3.2 故障树简化 .....	45-76
5.2 过盈联接的可靠性设计 .....	45-53	5.4 故障树的定性分析 .....	45-77
5.3 压缩螺旋弹簧的可靠性设计 .....	45-54	5.4.1 下行法 .....	45-78
5.4 滚动轴承的可靠性设计 .....	45-56	5.4.2 上行法 .....	45-78
5.4.1 滚动轴承寿命与可靠度之间的 关系 .....	45-56	5.5 故障树的定量分析 .....	45-79
5.4.2 滚动轴承额定动载荷与可靠度 之间的关系 .....	45-57	5.5.1 顶事件发生的概率 .....	45-79
		5.5.2 重要度 .....	45-80
		<b>第 5 章 维修性设计</b>	
<b>第 4 章 机械系统的可靠性</b>		1 概述 .....	45-82
1 不可修复系统的可靠性模型 .....	45-59	1.1 维修性的特征量 .....	45-82
1.1 串联系统 .....	45-59	1.2 维修的分类 .....	45-83
1.2 并联系统 .....	45-59	1.2.1 事后维修 .....	45-83
1.3 混联系统 .....	45-60	1.2.2 预防维修 .....	45-83
1.4 表决系统 .....	45-60	1.3 维修性设计与可靠性设计的 关系 .....	45-83
1.5 贮备系统的可靠度 .....	45-61	2 修复时间分布和维修度函数 .....	45-84
1.6 复杂系统 .....	45-62	2.1 修复时间分布 .....	45-84
2 可靠性预计 .....	45-63	2.2 维修度函数 .....	45-86
2.1 可靠性预计的目的 .....	45-63	2.2.1 已知修复时间为对数正态分布时 的维修度函数 .....	45-86
2.2 可靠性预计的方法 .....	45-63	2.2.2 已知修复时间为威布尔分布时的 维修度函数 .....	45-87
2.2.1 设计初期的概略预计法 .....	45-63	2.3 修复率 .....	45-87
2.2.2 数学模型法 .....	45-63	2.4 平均修复时间 .....	45-88
2.2.3 上下限法 .....	45-63	2.4.1 主动修复设备的平均时间 .....	45-88
2.2.4 蒙特卡罗模拟法 .....	45-64	2.4.2 平均维修工时(MMH) .....	45-89
3 可靠性分配 .....	45-65	3 按龄期更换的预防维修 .....	45-89
3.1 可靠性分配的原则 .....	45-65	3.1 按龄期预防维修设备的可靠度及 平均寿命 .....	45-89
3.2 可靠性分配的方法 .....	45-65	3.2 失效时间为威布尔分布时的预防 维修 .....	45-90
3.2.1 等分配法 .....	45-65	3.3 失效时间为指数分布时的预防 维修 .....	45-92
3.2.2 再分配法 .....	45-66	4 有贮备的可维修系统的可靠度 .....	45-93
3.2.3 比例分配法 .....	45-66	4.1 确定可维修系统可靠度的步骤 .....	45-93
3.2.4 综合评分分配法 (AGREE) .....	45-68	4.2 两单元并联系统的可靠度 .....	45-93
3.2.5 动态规划分配法 .....	45-68	4.2.1 系统的状态分析法 .....	45-93
4 失效模式、效应及危害度分析 (FMECA) .....	45-70	4.2.2 马尔可夫图解法 .....	45-95
4.1 基本概念 .....	45-70	5 可维修系统的有效度 .....	45-96
4.2 分析的过程与方法 .....	45-70		
4.3 FMECA 应用例子 .....	45-71		
5 故障树分析(FTA) .....	45-74		

5.1 有效度的定义 .....	45-96	3 基本统计分析 .....	45-112
5.2 可维修的单个部件的有效度 .....	45-96	3.1 描述性统计分析 .....	45-112
5.3 可维修的两单元并联系统的 有效度 .....	45-97	3.2 独立样本(成组)t-检验 .....	45-114
6 维修性设计例子 .....	45-98	3.2.1 独立样本 t-检验(1)(记录/个案格 式) .....	45-114
7 可靠性、维修性费用的计算模型 .....	45-102	3.2.2 相依(相关)样本 t 检验 .....	45-116
<b>第 6 章 可靠性分析中的统计分析软件应用</b>			
1 概述 .....	45-105	4 频数表分析 .....	45-118
1.1 运行环境 .....	45-105	5 概率分布计算 .....	45-120
1.2 基本功能 .....	45-105	5.1 贝塔(Beta)分布 .....	45-121
1.3 启动与退出 .....	45-106	5.2 Z(Normal)分布 .....	45-121
1.3.1 启动 .....	45-106	5.3 t(Student)分布 .....	45-122
1.3.2 退出 .....	45-106	6 非参数统计检验/分布拟合 .....	45-123
2 数据文件的建立 .....	45-106	6.1 常用描述性统计分析 .....	45-123
2.1 统计数据类型 .....	45-106	6.2 观察与期望频数的对照检验 .....	45-124
2.2 STATISTICA/Win 的文件类型 .....	45-107	6.3 柯尔莫哥罗夫-斯米尔诺夫两样本 检验 .....	45-125
2.3 建立数据文件 .....	45-107	6.4 分布拟合 .....	45-126
2.3.1 数据/大数据文件管理 .....	45-107	6.4.1 拟合连续型分布 .....	45-126
2.3.2 建立新的数据文件 .....	45-108	6.4.2 拟合离散型分布 .....	45-131
2.3.3 数据文件个案排序 .....	45-111	6.4.3 概率坐标纸 .....	45-134
		参考文献 .....	45-137

## 第 46 篇 实验设计法

### 第 1 章 实验设计与正交实验表

1 实验设计的作用 .....	46-3
1.1 实验设计的概念 .....	46-3
1.2 实验设计的领域 .....	46-3
1.3 实验顺序的随机性 .....	46-3
2 实验设计中因素和水平的确定方法 .....	46-3
2.1 因素的分类 .....	46-4
2.2 因素间相互关系 .....	46-4
2.3 多因素情况的处理 .....	46-5
2.4 水平的处理 .....	46-5
2.5 各种辅助方法 .....	46-5
3 正交实验表 .....	46-5
3.1 正交实验表的设计 .....	46-5
3.2 常用正交实验表 .....	46-6
3.3 方差分析和显著因素 .....	46-6
3.4 最优实验条件和平均值估计 .....	46-6
3.5 客车轮胎寿命实验 .....	46-6
3.6 喷淋洗涤实验 .....	46-7

3.6.1 配列与数据 .....	46-7
3.6.2 方差分析 .....	46-8
3.6.3 显著因素估计 .....	46-9
3.6.4 最宜条件的决定与预测 .....	46-9
3.7 铸铝机部件的实验 .....	46-9
3.7.1 配列与数据 .....	46-9
3.7.2 方差分析 .....	46-10

### 第 2 章 基本实验设计方法

1 裂区法设计 .....	46-13
1.1 单元裂区设计和因素分类 .....	46-13
1.2 正交表的列的分组 .....	46-13
1.3 薄板实验 .....	46-14
1.3.1 配列与数据 .....	46-14
1.3.2 方差分析 .....	46-14
1.3.3 因素的估计与最宜条件 .....	46-15
1.4 三水平系统的裂区设计 .....	46-17
1.4.1 因素与水平 .....	46-17
1.4.2 配列与数据 .....	46-17





3.2 因素与水平 .....	46-56	8.3 配列与实验 .....	46-72
3.3 配列与数据 .....	46-57	8.4 数据 .....	46-72
3.4 数据分析 .....	46-58	8.5 累积法分析 .....	46-74
3.5 结论 .....	46-59	9 电话交换机的颜色调整实验 .....	46-75
4 羊毛纺纱中的清洗过程和梳理过程		9.1 目的 .....	46-75
实验 .....	46-60	9.2 因素与水平 .....	46-75
4.1 目的 .....	46-60	9.3 配列 .....	46-76
4.2 因素与水平 .....	46-60	9.4 方差分析与估计 .....	46-77
4.3 数据与配列 .....	46-61		
4.4 数据分析 .....	46-61	<b>第4章 测量技术的实验设计方法</b>	
4.4.1 一般的方差分析 .....	46-61	1 校正方法与测量误差 .....	46-78
4.4.2 分散指数法 .....	46-61	2 比例校正与SN比 .....	46-78
4.4.3 变数变换法 .....	46-61	3 测量误差及其改善方法 .....	46-78
4.5 结论 .....	46-62	4 校正频率与误差变动 .....	46-79
5 聚酯纤维混纺布料的生产条件 .....	46-63	5 信号因素水平的确定方法 .....	46-80
5.1 因素与水平 .....	46-63	5.1 化学分析的例子 .....	46-80
5.2 配列与数据 .....	46-63	5.2 质量测定的例子 .....	46-81
5.3 方差分析 .....	46-64	5.3 以轮圈动平衡测量为例子 .....	46-81
5.3.1 计算 .....	46-64	6 误差因素的选取方法 .....	46-83
5.3.2 估计 .....	46-65	7 改善SN比的实验设计 .....	46-84
5.3.3 最宜条件与工序平均 .....	46-66	7.1 信号因素与误差因素 .....	46-84
6 密封工序和灯泡外观品质的实验 .....	46-66	7.2 SN比的计算 .....	46-84
6.1 目的 .....	46-66	7.3 误差因素及其控制措施 .....	46-84
6.2 因素与水平 .....	46-67	7.3.1 内部正交表——控制因素的	
6.3 配列与实验结果 .....	46-67	安排 .....	46-84
6.3.1 累计法分析,方差分析 .....	46-67	7.3.2 外部正交表——信号因素的	
6.3.2 累计法,因素效应估计、最宜		安排 .....	46-84
条件的决定 .....	46-68	8 求解标准固有误差的实验设计方法 .....	46-84
6.4 说明 .....	46-68	8.1 标准的误差 .....	46-84
7 电灯泡的瓦数,功率和寿命实验 .....	46-68	8.2 直读天平的实验 .....	46-85
7.1 目的 .....	46-68	8.3 方差分析 .....	46-85
7.2 因素与水平 .....	46-68	8.4 标示值误差的问题与数据分析 .....	46-86
7.3 配列与数据 .....	46-69	9 标准稳定性的实验设计与数据分析 .....	46-86
7.4 数据分析 .....	46-69	9.1 标准稳定性 .....	46-86
7.5 估计 .....	46-70	9.2 稳定性的实验设计 .....	46-86
7.6 结论 .....	46-71	9.3 数据分析 .....	46-87
8 氧化膜实验 .....	46-71	9.4 变化量的预测 .....	46-87
8.1 目的 .....	46-71	<b>参考文献</b> .....	46-88
8.2 因素与水平 .....	46-71		



3.2	机构的并联式组合	47-27
3.3	机构的叠合式(或运载式)组合	47-27
3.4	机构的叠联式组合	47-28
3.5	组合机构	47-28
4	广义机构	47-29
4.1	液动机构	47-29
4.2	气动机构	47-30
4.3	电磁机构	47-30
4.4	振动机构	47-30
4.5	光电机构	47-31
5	执行机构的创新方法	47-32
5.1	应用机构学原理	47-32
5.2	利用连杆机构或连架杆机构的运动特点构思新的机构	47-33
5.3	用成型固定构件和相对运动实现复杂运动过程	47-35
5.4	利用多种驱动原理创新机构	47-35
5.5	机构类型创新和变异设计	47-37
5.6	机构类型变换法	47-38
6	机构选型	47-39
6.1	按运动形式选择机构	47-39
6.2	按执行机构的功用选择机构	47-41
6.3	按不同的动力源形式选择机构	47-41
6.4	按先易后难选择机构	47-41
6.5	选择机构及其组合安排时应考虑的主要要求和条件	47-41
7	动作解法库的建立	47-41

## 第4章 机械运动系统的协调设计

1	机械运动系统的基本构成	47-42
1.1	传动系统	47-42
1.2	执行系统	47-42
2	机械运动系统设计	47-42
2.1	机械运动系统的基本内容	47-42
2.2	机械运动系统的集成设计	47-47
3	执行机构的协调设计	47-49
3.1	机器的机构传动系统类型和工作原理	47-49
3.2	机器执行机构的协调设计	47-51
3.3	执行机构协调设计的分析计算	47-52
4	机械运动循环图设计	47-53
4.1	机器的运动分类	47-53
4.2	机械的运动循环周期	47-53
4.3	机器的工作循环图	47-53
4.4	拟定机器工作循环图的步骤和方法	47-55

4.5	机器工作循环图的作用	47-55
4.6	机械运动循环图设计举例	47-55

## 第5章 机械系统运动方案的构思和设计

1	机械系统运动方案设计的主要步骤和内容	47-58
1.1	机械系统运动方案设计的主要步骤	47-58
1.2	功能原理方案设计	47-58
1.3	机械系统运动方案设计	47-59
1.4	机械系统运动方案的尺度综合	47-60
2	机械的工艺动作过程的构思	47-60
2.1	工艺动作过程是功能和功能原理方案的具体体现	47-60
2.2	工艺动作过程与机器类型的关系	47-62
2.3	工艺动作过程构想原则	47-62
2.4	工艺动作过程的构思方法	47-63
3	机械工艺动作过程分解和执行机构的选择	47-64
3.1	机械工艺动作过程的内涵	47-64
3.2	机械工艺动作过程的分解	47-65
3.3	动作组合的创新	47-67
3.4	动作的描述和机构属性表达方式分析	47-69
3.5	执行机构的选择	47-70
4	机械运动系统方案的组成原理与方法	47-71
4.1	机械运动系统组成的相容性原则	47-71
4.2	机械运动系统组成的系统最优化原则	47-72
4.3	寻求执行机构的创新设计是机械运动系统创新设计的基础	47-72
5	机械系统运动方案设计举例	47-72
5.1	设计平版印刷机的运动系统方案	47-72
5.2	设计冲压式蜂窝煤成型机的运动系统方案	47-75

## 第6章 机电一体化系统概念设计

1	概述	47-79
1.1	机电一体化系统的形成和发展	47-79
1.2	机电一体化系统的定义	47-79
1.3	机电一体化系统的研究状况	47-79
1.4	机电一体化系统的组成	47-80
1.5	机电一体化系统组成的新认识	47-81
2	机电一体化系统的应用和特点	47-81
2.1	机电一体化系统的应用	47-81
2.2	机电一体化产品的主要特点	47-82

3 机电一体化系统设计过程模型及数学描述 .....	47-82	第 7 章 机械运动系统的评价体系和评价方法	
3.1 机电一体化系统设计过程模型的建立 .....	47-82	1 评价指标体系的确定原则 .....	47-98
3.2 机电一体化系统设计过程的数学描述 .....	47-84	2 评价指标体系 .....	47-98
4 广义执行机构子系统的类型和设计 .....	47-84	2.1 机构的评价指标 .....	47-98
4.1 传统执行机构 .....	47-84	2.2 几种典型机构的评价指标的初步评定 .....	47-99
4.2 广义执行机构 .....	47-84	2.3 机构选型的评价体系 .....	47-99
4.3 广义执行机构的种类和基本特性 .....	47-85	2.4 机构评价指标的评价量化 .....	47-99
4.4 驱动元件的机械特性的基本特点 .....	47-86	2.5 机构系统选型的评估方法 .....	47-100
4.5 驱动元件与执行机构的匹配 .....	47-88	3 价值工程方法 .....	47-100
4.6 广义执行机构的运动方程式 .....	47-88	3.1 产品的功能 .....	47-100
5 检测传感子系统类型和设计 .....	47-88	3.2 产品的寿命周期成本 .....	47-100
5.1 检测传感器的分类与基本要求 .....	47-88	3.3 产品的价值 .....	47-100
5.2 位移检测传感器 .....	47-89	3.4 机械运动方案的价值评定 .....	47-101
5.3 速度、加速度传感器 .....	47-90	4 系统分析方法 .....	47-101
5.4 力、力矩传感器和其他传感器 .....	47-90	4.1 系统工程评价方法的基本原则 .....	47-102
5.5 传感器与微机的接口 .....	47-90	4.2 建立评价指标体系和确定评价指标值 .....	47-102
5.6 检测传感系统的设计原则 .....	47-91	4.3 建立评价模型 .....	47-102
6 信息处理及控制子系统的设计 .....	47-91	5 模糊综合评价法 .....	47-103
6.1 信息处理及控制子系统的基本构成 .....	47-91	5.1 模糊综合评价中主要运算符号 .....	47-103
6.2 信息处理及控制子系统的一般设计过程及内容 .....	47-92	5.2 模糊集合的概念 .....	47-103
6.3 微型计算机的种类 .....	47-92	5.3 隶属度函数的确定方法 .....	47-104
6.4 主要控制方法 .....	47-93	5.4 模糊综合评价 .....	47-104
6.5 控制系统的种类 .....	47-94	6 实例分析 .....	47-107
6.6 动作控制方式及其特点 .....	47-95	6.1 系统工程评价法评价机械运动方案 .....	47-107
7 机电一体化系统设计示例 .....	47-96	6.2 模糊综合评价法评价机械运动系统方案 .....	47-109
7.1 线料自动切断机 .....	47-96	参考文献 .....	47-113
7.2 多功能缝纫机的横针机构 .....	47-96		

## 第 48 篇 反求设计与快速成形制造技术

### 第 1 章 反求工程技术

1 反求工程的概念和流程 .....	48-3	2.3 曲面重构技术 .....	48-16
1.1 广义反求工程 .....	48-3	2.4 实体构造技术 .....	48-22
1.2 制造业反求工程的一般流程 .....	48-3	3 反求工程的应用领域 .....	48-22
2 反求工程所涉及的主要技术 .....	48-4	4 反求测量系统简介 .....	48-23
2.1 数据采集技术 .....	48-4	4.1 光学测量设备介绍 .....	48-24
2.2 数据处理技术 .....	48-5	4.2 重构软件介绍 .....	48-26
		5 应用实例 .....	48-28
		6 反求设计全过程实例 .....	48-29

<b>第 2 章 快速成形</b>		5 快速成形相关软件技术 .....	48-37
1 快速成形设备发展简况 .....	48-31	6 主要的快速成形产品介绍 .....	48-38
2 快速原型技术流程 .....	48-31	7 快速成形技术的应用实例 .....	48-40
3 几种典型的快速成形技术介绍 .....	48-32	8 快速成形技术的发展方向 .....	48-40
4 快速成形技术的应用 .....	48-33	<b>参考文献</b> .....	48-41

## 第 49 篇 互联网上的合作设计

<b>第 1 章 互联网上合作设计的意义</b>		<b>第 4 章 互联网上合作设计的支撑技术</b>	
1 现代设计一般过程的描述 .....	49-3	1 群体合作技术 .....	49-15
1.1 需求的确认 .....	49-3	1.1 CSCW 研究的发展 .....	49-15
1.2 技术可能扫描 .....	49-4	1.2 CSCW 研究的内涵 .....	49-15
1.3 概念设计 .....	49-1	1.3 CSCW 和群件的关系 .....	49-15
1.4 技术经济分析 .....	49-4	2 产品设计信息共享技术 .....	49-16
1.5 详细设计 .....	49-4	2.1 STEP 技术 .....	49-16
2 现代设计的基本特征 .....	49-4	2.2 XML 技术 .....	49-17
3 设计为什么要在网上合作 .....	49-5	3 设计知识资源的构建、发布、发现和集成技术 .....	49-17
<b>第 2 章 设计知识服务和分布式智力资源</b>		3.1 TCP/IP 协议系列 .....	49-17
1 设计中的知识 .....	49-7	3.2 分布式对象技术 .....	49-18
1.1 产生知识的信息源分类 .....	49-7	3.3 Web Services 技术 .....	49-19
1.2 设计知识的结构特征 .....	49-7	3.4 UDDI 技术 .....	49-20
2 获取信息的资源 .....	49-8	3.5 智能体 Agent .....	49-22
2.1 虚拟现实需要的资源 .....	49-8	3.5.1 Agent 的基本概念 .....	49-22
2.2 物理模型试验需要的资源 .....	49-9	3.5.2 Agent 的属性 .....	49-23
2.3 样机试验需要的资源 .....	49-9	3.5.3 Agent 的优点、局限性和面临的挑战 .....	49-23
2.4 在产品运行状态监测需要的资源 .....	49-9	4 设计过程管理技术 .....	49-24
2.5 其他信息资源 .....	49-10	4.1 PDM 技术 .....	49-24
2.5.1 已有知识 .....	49-10	4.2 安全控制 .....	49-24
2.5.2 市场信息 .....	49-11	<b>附录 知识资源 UDDI 注册信息填写表</b> .....	49-25
<b>第 3 章 分布式智力资源的运作模式</b>		<b>第 5 章 现代设计与制造网上合作研究中心及相关的资源</b>	
1 智力资源的构成——服务提供方 .....	49-12	1 中心的创建与进展 .....	49-27
1.1 智力资源的构成要素 .....	49-12	1.1 中心的创建 .....	49-27
1.2 智力资源的生存条件 .....	49-12	1.2 中心的进展 .....	49-27
2 设计实体(服务请求方)的构成要素 .....	49-13	2 中心的网上资源介绍 .....	49-28
3 合作设计的层次结构 .....	49-13	2.1 性能分析评估服务 .....	49-28
		2.2 支持设计的数据库服务 .....	49-29

2.2.1 系统的主要功能 .....	49-29
2.2.2 系统的特点 .....	49-30
2.3 性能实验评估服务 .....	49-30
2.4 服务供应商的评估服务 .....	49-31
2.4.1 供应商信息管理 .....	49-31
2.4.2 指标体系管理 .....	49-31
2.4.3 供应商评价 .....	49-31
2.5 虚拟仪器服务 .....	49-32
2.5.1 背景及意义 .....	49-32
2.5.2 功能介绍 .....	49-32
3 中心资源的使用导航 .....	49-33
3.1 如何使用智力资源的统一发布、发现和集成(UDDI) .....	49-33
3.1.1 用户注册 .....	49-33
3.1.2 发布知识服务实体和所提供的服务的信息 .....	49-33
3.1.3 搜索知识服务实体和所提供的服务的信息 .....	49-34
3.2 如何组织远程会议 .....	49-34
3.2.1 远程会议实现背景 .....	49-34
3.2.2 远程假体异地合作设计的业务流程 .....	49-34
3.2.3 远程会议实现的环境 .....	49-34
4 中心的发展方向 .....	49-35

附录 虚拟研究开发中心合同 .....	49-35
---------------------	-------

## 第 6 章 互联网上的合作设计实例

1 项目背景 .....	49-38
2 涡轮膨胀机采用动压滑动轴承支承的缺点 .....	49-39
3 涡轮膨胀机采用主动电磁轴承支承的优点 .....	49-39
4 互联网上的合作设计过程 .....	49-39
4.1 知识资源注册 .....	49-39
4.2 搜索设计资源单元并评估 .....	49-40
4.3 初步组成虚拟设计联盟 .....	49-40
4.4 主动电磁轴承的结构设计 .....	49-40
4.5 转子轴承系统的动力学分析 .....	49-41
4.6 涂层设计 .....	49-41
4.7 与厂家交换设计意见 .....	49-42
4.8 可铸造性评估 .....	49-42
4.9 可加工性评估 .....	49-43
4.10 可装配性评估 .....	49-43
4.11 制造 .....	49-43
4.12 基于 Internet 的远程试验 .....	49-43
4.13 台架试验 .....	49-43
参考文献 .....	49-44

## 附录 现代设计主流软件

### 第 1 章 CAD/CAM 软件

1 UGII 软件 .....	3
2 Pro/Engineering 软件 .....	9
3 I DEAS 软件 .....	14
4 CATIA 软件 .....	20
5 AUTOCAD2004 .....	31
6 MDT 软件 .....	33
7 Solid Edge .....	36
8 MasterCAM 软件 .....	46
9 EdgeCAM 数控自动编程系统 .....	48
10 CAXA 系列 CAD/CAM/CAE 软件 .....	49
11 金银花 CAD/CAM 系统 LONICERA .....	51
12 InteCAD .....	53
13 PICAD 2000 软件产品 .....	54
14 CIMATRON .....	55
15 ADCEUS .....	58

### 第 2 章 CAPP 软件

1 SIPM/CAPP .....	60
2 Inte/CAPP .....	64
3 开目 CAPP .....	65
4 高华 CAPP .....	65
5 XTCAPP .....	65

### 第 3 章 PDM 软件

1 SIPM/PDM .....	70
2 iMAN .....	74
3 Windchill .....	79
4 CATIA/PDM II 产品开发管理解决方案 .....	83
5 Inte PDM 系统 .....	85
6 “金银花”系列 BEE-PDM .....	86
7 CV/Optegra .....	88
8 ISS/InSync .....	88

9 SherpaWORK ..... 88

#### 第 4 章 计算机辅助工程分析(CAE)软件

1 ANSYS 软件 ..... 90

2 工程分析软件 NASTRAN ..... 93

3 机械系统动力分析软件 ADAMS ..... 94

4 结构有限元分析与优化设计软件系统  
JIFEX2.0 版 ..... 95

5 机械系统动力学、运动学分析软件

DADS ..... 98

6 MSC.DYTRAN 软件 ..... 102

7 铸造工艺分析软件 InteCAST 系统 ..... 105

8 动力学分析设计师 Dynamic Designer

/Motion ..... 107

参考文献 ..... 109



# 第36篇 创新设计

主 编 陈 铭  
编写人 陈 铭

# 第 1 章 创造与设计思维原理和方法

## 1 创新思维和方法

创造过程是直觉与逻辑思维方式交叉、综合的工作过程。

### 1.1 创新思维

#### (1) 直觉思维

思维可按表 36.1-1 分类。

在下意识状态下,对事物内在复杂关系突发式的领悟过程,具有创造灵感忽然降临的色彩。

表 36.1-1 思维类型

类型	性质	特点	影响因素
形象思维	使用反映同类事物一般特征的形象	头脑中浮现形象,类同形象引起联想和类比	右大脑的发达程度,右大脑通过胼胝体与左大脑交互作用的发达程度
抽象思维	使用反映事物或现象本质属性的概念和推理	从不同的具体问题中抽象出具有共性的概念来	左大脑的发达程度,左大脑通过胼胝体与右大脑交互作用的发达程度
发散性思维	根据问题,不按常规,沿着不同的正向、逆向、多向思维和角度,多方面寻求问题的各种可能解答	流畅:反应敏捷,在较短时间内想出多种方案;灵活:触类旁通,随机应变,不受心理定势影响;独特:所提出的解决方案有特色	人的知识广博程度,非本行多方面科技文化领域的涉猎,活化知识存储方式,使知识脱离特定情景、重新组合
收敛性思维	来自多方面的知识信息聚合于同一体	分析比较,得出从不同角度出发的优缺点;综合推理、收敛引出最优答案	细致的分析能力;严谨的逻辑推理能力
逻辑思维	思维过程中严格遵守逻辑规则	注意事物的显性质和常规功能,更要发掘事物的潜性质和非常规功能;以抽象概念为其思维元素;使用固定范畴和程式,把复杂问题化简,找出主要因素。若前提条件不完全,则结论带有或然性,不够新颖	把事物分解为不同部分、单元、要素的能力;综合、归纳、演绎推理的能力
非逻辑思维	思维过程中不严格遵循逻辑规则,灵活自由,凭想象、直觉和灵感	结论往往突破常规,新颖独创。或然性很大,用以启迪心智,扩展思路	联想力:由一事物引发想到另一事物,是想象、直觉、灵感的基础。想象力:加工改造,产生新形象。直觉:不受固定逻辑约束,突然自动地直接领悟事物本质,成果突破常规,但不一定可靠。灵感:偶然机遇使人有着迷于问题时的全部积极心理活动突然连锁激发,不能控制的潜意识进入显意识,爆发出创造火花
直达思维	思考问题始终不离开问题的情境和要求	直接面对问题的约束条件,解决简单问题特别有效	问题的性质和人的能力
旁通思维	通过分析把问题转换成另一领域的等价问题	全面细致地分析问题的情境和要求,把问题转换一个角度重新表达	分析问题抓住本质的能力,对类比、模拟、仿生、移植、换元等转换问题方法的掌握

直觉式创造方式具有如下缺点:

1) 很少能在需要的时刻获得正确的、突降的灵感妙想;

2) 其结果在极大程度上取决于处理者的天赋和经验;

3) 所得的解很可能被其自身的业务水平所约束。

### (2) 逻辑思维

根据所提出的创造目标,进行逻辑推理式的思维,把目标展开、分解和综合,寻求各层分目标的解和最终整体解,用主动的、可按步就班的工作方式向目标逼近。

### (3) 创造过程

提出一个创造目标,这个目标本身也可能就是一个创造灵感;为了实现这个目标,必须进行分析 and 逻辑推理;这个过程中出现的技术难关,必须经过反复试验才能找到最终解;而这些解法的发现,往往又带有突发性、直觉性,最终实现创造目标。

## 1.2 创新思维机制

### (1) 大脑皮质

按其结构可分为新皮质和旧皮质。新皮质的功能包括认识、判断和意志等高级精神活动以及由此产生的行动。旧皮质的功能是管理人的本能的欲望和愉快、烦恼等情绪。人的一切有意识的思想行动由新皮质控制,一切下意识的思想行动都由旧皮质控制。

### (2) 输入大脑的信息

一种是有意识记忆的信息,存储在新皮质中,其逻辑性强,有时间序列关系,质量较高;另一种是附带的、下意识记忆下来的信息,存储在旧皮质中。存储在新皮质中的信息如果长期不用,也会随着脑细胞的新陈代谢,解体为个体的、分散的信息,转入旧皮质的深处,不易通过新皮质的意识活动加以调用。

### (3) 信息调用过程

逻辑思维在新皮质中进行,通过新旧皮质间的联系通道,有步骤地、逻辑地调用旧皮质中的知识信息,逻辑思维过程即冥思苦想、艰苦探索的过程。这一过程中,由于偶然因素作用,使旧皮质中的有用信息突发性地调出,使原先的逼近过程突发到位,形成逻辑思维与直觉思维交叉的现象。

### (4) 量变质变过程

创新灵感的产生过程可理解为从量变到质变的转化过程。量变的第一含意是信息量的增加。创造的过程通常伴随着艰苦的学习和实践,积累了足够的知识信息,构成了创新思维的基础。量变的第二含意是联系量的扩大。通常是一个艰苦的创造思维过程,通过逻辑推理,不断缩小对关键信息的包围网,最终形成创新灵感。

## 1.3 创新的外部因素与思维法则

### (1) 培育创新思维的外部因素

1) 构造激励创新的环境 只有在激情的支配下,人脑的智力活动才能被高度激发,甚至形成突发灵感。创造激情可能由市场压力或诱惑力等经济因素促成,也可能出自科学家的崇高事业心,以及政治需要等。激励的手段可以是精神鼓励、经济刺激、政治动员,甚至民族感情。

2) 强化信息获取的力度 创造是建立在大量知识和信息的基础上,信息的快速获取将成为成败的关键因素。

3) 创造知识融合的条件 创造发明往往是多学科知识融合的结果。例如鼓励边缘学科的建立,强化多学科间的学术交流等。

### (2) 创新思维法则

#### 1) 分析与综合思维法则

分析 面对所提出的创新对象、任务或设想,把本质和非本质的问题分解为逻辑层次与关系清晰的各级子问题。

综合 在求得各子问题的解之后,通过建立联系,实现最终需要的总体功能。综合过程常常是一个创新过程。

#### 2) 收敛与发散思维法则

收敛式思维 确定一个思维焦点,即创造目标,寻求可能的通向这个目标的途径,最终得到一条最佳途径,它使目标成为可实现的解,并且与理想解的差异是最能够被接受的。

发散式思维 以一个原始解为出发点,寻找更多更好的解。这些新解不一定是原始解的变异,它很可能是在保证系统功能的前提下,一种新的原理和系统。

要点 对目标或原始解的功能抽象化,将所希望创造的对象特征用最本质、最核心的关系词在高层次上加以描述,使这种关系具有最大的通用性,可以更好地激发创新思维的进行。

例如,要创造一种“钉在水泥墙面上的钉子,用它悬挂物品”如此具体的任务描述只能引导人们创造出水泥墙面钉。如果把创造对象描述为“置于垂直坚硬面上的悬挂手段”,就把钉子这一预设解去除了。人们在这抽象化的目标下,可以创造出胶粘钩、吸附钩、带膨胀螺钉的挂钩等更多种解。

3) 对应与联想思维法则 通常对应联想法则具有如下4种类型:

① 具象类比 把现有的创造任务或意想中的系统转移到具体类比物上去,把类比系统作为意想系统的模型加以考查分析,从中悟出问题的解。

例如,用两栖动物的生理特点来类比水陆两用交通工具,用在夜间活动的动物的视觉机制来类比夜视装置等等。

② 抽象类比 用能抽象反映问题概念的词或简练词组来类比问题,表达所探讨问题的关键,然后以这些词汇为出发点,展开思维。

例如,人们想发明一种开罐头的工具,于是选择“开”为关键词,由此出发,产生撬开、剥开、撕开、拧开、揭开、破开、剖开、划开等联想,最后在剖开、划开中找到可行方案,创造出开罐刀,并随着罐头密封方式的变化,撬开、拧开也成为对应动作,创造出相应的工具。

③ 拟人类比 即把自己想象为所讨论问题中的某个因素,设身处地地投入所讨论的工作环境中,想象自己会遇到的问题和感受,从中领悟问题的解。

如设计一台洗衣机,并认为增强去污能力是创新的关键,把自己想象成污垢,紧紧地粘到了纤维上,研究“洗衣机该拿我怎么办?”。于是可想象,“如果水中有化学物质,把我和纤维的粘连溶解开了,我就成为自由粒子;洗涤动作都在我头顶上走过,对我毫无作用,如果有股纤细而有力的水流不断从纤维网隙中冲过,我将被带出缝隙;最好这股水流把我冲出后随即把我带走,千万不要冲进冲出,使我老在网缝的附近转,更不要出了一个网缝又进到另一个网缝;当我脱离了网缝后,尽管随水流运动,但始终跑不出去怎么办?因为水很自由,什么缝隙都钻得出去,但我太大,一路阻拦很多,布料表面好象还有许多纤维小手,一个个拉住我不放。”人们可以如此浮想联翩,直到想象自己进入洗涤剂的水池中,由清水带离为止,而这一想象过程同时伴随着发现问题和寻求解的过程。

④ 幻想类比 幻想常常寓于神话故事中,但随着技术的进步,许多已变为现实。今天,采用科技知识对未来作出幻想,如上月亮居住、到海底中生活等等。

#### 4) 分解与组合思维法则

分解:

① 事物本身由很多可分解的功能组成,组成方式不同,总体功能将不同,分解是为了重组;

如:将机床分解为若干标准模块,成为可按用户需求组合的组合机床。

② 事物的主要功能常和一些辅助功能搭配在一起,将主功能分解出来,使其有可能获得更好的发挥,或组成更多更好的主辅搭配;

如:用标准电子钟芯组成各种工艺钟。

③ 事物由多种功能组成,合在一起彼此影响,分离有利于各自特点的发挥,或有利于创新改进。

如:分体式空调、母子电话机等。

组合:

① 辅助功能的组合

如:轿车的各种选装设备。

② 同种功能的组合

如:双体船、双管猎枪、双筒望远镜等。

③ 无关功能的组合

如:车与船的组合创造出水陆两栖坦克、鱼雷和飞机的组合创造出鱼雷攻击机。

#### 5) 换元与移植思维法则

换元:

客观事物内在的运动规律虽然表现的形式不同,但规律相同,数学很好地反映了这一点。这种思维方法转移到创造学中,成为换元法则。

如:美国核物理学家格拉塞尔喝啤酒时,看到啤酒杯中一串串上升的气泡,猛然想到自己一直在研究的课题——怎样探测高能粒子飞行轨迹。于是,他就用啤酒代替高能粒子穿越的介质,顺手捡起几粒碎小鸡骨代替高能粒子,等到酒杯中气泡冒完后,将其丢入杯中啤酒里,只见随着碎骨粒的沉落,周围不断冒出气泡,用气泡显示出碎骨粒下降过程的轨迹。这里,啤酒杯里揭露出一种方法和规律,即可以利用粒子经过后在介质中留下的滞后反应显示粒子的运动轨迹。经过不断试验,当带电粒子穿过液态氢时,所经路线同样出现了一串串的气泡,终于以这种方法清晰地呈现出粒子飞行的轨迹。在这个例子里,鸡骨粒在啤酒中运动和电子在液态氢中运动,彼此具有换元关系。

移植:

某一领域内的一种方法或研究成果在另一领域内的应用,它常导致一门新的学科、领域或学说的创立。

如:将力学移植到生物学中,产生生物力学;将军事技术用于民用工业,产生许多民用新技术,等等。

#### 6) 正向、迂回与反向思维法则

正向、迂回与反向思维是人们常用的思维方式。不同的思维进程都可能到达创新的终点,关键在于灵活运用。

① 正向思维 是人们常用的、朝向目标直线前进的逻辑思维进程。

② 迂回思维 是在直线思维之外,通过间接、旁路的思维进程达到目标。

如:在快速原型技术中,通过制造平面实现三维实体的制造。

③ 反向思维 又称逆向思维,是创新思维中重要的思维方式,离经叛道,不为现有的思维定势所束缚,是创新人才的特点。因为所谓公认的思维方向常常与目标反向。

如:人们在环线上驱车前进,需很长时间,如调转

(续)

车向,很快就能达到目标,因目标实际上就在紧贴出发点的背后;又如:几何学中的反证法。

### 1.4 创造活动的组织科学

(1) 创造技法基本原理 (见表 36.1-2)

表 36.1-2 创造技法基本原理

主动原理	创造者积极主动设问探索,保持强烈的创造冲动
刺激原理	创造者善于吸纳信息,对新鲜事物有强烈兴趣并进行追踪
希望原理	不满足现状和已得经验,怀疑提问,追求事物的完善化、理想化

环境原理	有容许探索、从失败中求成功的社会环境,使创造者保持良好心境
量多原理	解决任何问题都有尽可能多的创造性设想和方案,多多益善,没有数量就没有质量
群众原理	人人都有创造力,群众智慧大于个人智慧
激发原理	创造性思维相互激励、活化、引伸、改进、综合,能引发出更多的新颖设想
自由原理	打破定势,不受权威和传统逻辑或常规限制,敢于提出被一般人认为荒唐的看法

(2) 创造技法类型 (见表 36.1-3)

表 36.1-3 创造技法类型

智力激励法	步骤	<ol style="list-style-type: none"> <li>1. 明确会议中心议题,以简单为好,复杂问题要化为多个单一问题分别讨论</li> <li>2. 确定会议人选,5~15人为宜,多数是该议题领域的内行,少数是知识广博的外行</li> <li>3. 坚持会议规则,自由发言,禁止评判。互相启发,追求数量</li> <li>4. 与会者思维发散,畅谈新奇设想,一般不超过1h</li> <li>5. 会后对设想进行整理评价,评价人员一般5人为宜,评价指标包含科学与技术、生产与市场(社会)</li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 信息直接传递,相互激励强度大,造成创造环境</li> <li>2. 会议易受外向型性格的人控制,内向型不易发挥</li> </ol>
群体法	635法 步骤	<ol style="list-style-type: none"> <li>1. 6人与会,明确议题</li> <li>2. 每人在卡片上写出3种设想方案,5min为一单元,卡片相互交流</li> <li>3. 第二个5min单元中,每人根据相互启发,再在卡片上写出3种设想</li> <li>4. 如此循环,半小时可得108种方案</li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 书面交流</li> <li>2. 每人意见都得到反映</li> <li>3. 思维冲撞程度差些</li> </ol>
集智法	CBS法 步骤	<ol style="list-style-type: none"> <li>1. 明确会议主题</li> <li>2. 邀3~8人与会</li> <li>3. 每人发50张卡片,每张填一设想,共5min</li> <li>4. 轮流宣读卡片,每次读一张后别人质询,有新启发写在新卡片上,共30min</li> <li>5. 交流讨论20min,共1h</li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 书面与口头结合</li> <li>2. 质询、改善,筛选纳入议程</li> </ol>
Ddphi法	步骤	<ol style="list-style-type: none"> <li>1. 组织者针对问题,编写意见征询表寄有关专家</li> <li>2. 专家在限期内填好寄回</li> <li>3. 收到复函后,概括、整理、综合后的意见和征询表再寄有关专家</li> <li>4. 反复多次</li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 专家开始独立思考,随后交流激励</li> <li>2. 时间较充裕,设想较实用可靠</li> <li>3. 时间较长</li> <li>4. 缺少激励环境气氛</li> </ol>

(续)

系 统 探 求 法	5W2H 法	步 骤	<p>例：机械产品设计</p> <ol style="list-style-type: none"> <li>1. Why? 为何设计该产品，采用该结构？</li> <li>2. What? 产品有何功能？需否创新？</li> <li>3. Who? 产品用户是谁？谁来设计？</li> <li>4. When? 何时完成该设计？各阶段如何划分？</li> <li>5. Where? 产品用于何处？在何处生产？</li> <li>6. How to do? 结构如何设计？形状？材料？</li> <li>7. How much? 单件还是批量生产？</li> </ol>
		特 点	<ol style="list-style-type: none"> <li>1. 适于任何工作，只是发问具体内容不同</li> <li>2. 可突出其中任何一问</li> </ol>
	奥 斯 本 设 问 法	步 骤	<p>例：机械产品设计检核表</p> <ol style="list-style-type: none"> <li>1. 转化：该产品能否稍作改动或不作改动后移作他用？</li> <li>2. 引伸：能否从该产品中引出其他产品？或用其他产品模仿该产品？</li> <li>3. 变动：能否对产品进行某些改变？结构？运动？造型？工艺？……</li> <li>4. 放大：该产品放大（加厚，加深……）后，能否改变其性能？</li> <li>5. 缩小：该产品缩小（减薄，变浅……）后如何？</li> <li>6. 颠倒：能否正反（上下、前后）颠倒使用？</li> <li>7. 替代：该产品能否用其他产品替代？</li> <li>8. 重组：零件能否更换？</li> <li>9. 组合：现有几个产品能否组合为一个产品？或零部件组合？或功能组合？……</li> </ol>
		特 点	<ol style="list-style-type: none"> <li>1. 由不同角度提问题</li> <li>2. 把不同角度列成检核表逐一检查</li> <li>3. 检核表可补充扩展，例如操作性扩展：增加、分割、缓和、代换、展开、象征等</li> <li>4. 检核表可变型为针对更具体问题的检核表，例如对产品设计过程：增加功能，提高性能，降低成本，增加销售等</li> </ol>
联 想 类 比 法	联 想 发 明 法	特 性 列 举 法	<ol style="list-style-type: none"> <li>1. 选择需要改进的对象</li> <li>2. 列表分解该对象的组成部分：名词性特征（部件、材料等）、形容词特征（性质、形状等）、动词性特征（功能、作用等）</li> <li>3. 编制该对象的组成部分的本质特征表</li> <li>4. 分析所有特征，用取代、替换、简化、组合等方法加以改进</li> <li>5. 重新设计</li> </ol>
		特 点	<ol style="list-style-type: none"> <li>1. 规范化：按一定规范进行，而不是随机列举</li> <li>2. 全面性：将对象所有特性都列举出来，系统地思考解决办法</li> <li>3. 特性列举法可变型为其他列举法，如缺点列举法：分析缺点及其存在条件，在原条件下改缺点为优点；或改换原条件使原缺点成为优点。又如希望列举法：从市场用户意愿出发，提出创新设想</li> </ol>
		步 骤	<ol style="list-style-type: none"> <li>1. 无约束地由一事物自由联想到另一事物，可能产生一些新联系，激发新思路</li> <li>2. 将某些看似无关的事物，通过插入若干中间事物联想在一起</li> <li>3. 把待解决事物的边界条件加入，进行控制联想。方法有：由一事物联想到在空间或时间上与其接通的另一事物；由一事物联想到与其有类似特点的另一事物（功能、性质、结构等）；由一事物联想到与其对立的另一事物；由一事物联想到与其有因果关系的另一事物；由一事物联想到与其有从属关系的另一事物</li> </ol>
		特 点	<ol style="list-style-type: none"> <li>1. 头脑中储存的信息要多，知识面要广</li> <li>2. 参加实践，熟悉客观实际的发展变化</li> <li>3. 要求掌握联想规律</li> <li>4. 联想发明法可变型为其他联想法，如强制联想法，把几种一般看来无关的产品强行联合，构成创新产品</li> </ol>

(续)

联想类比法	步骤	<ol style="list-style-type: none"> <li>1. 把不同知识领域的专家组合在一起</li> <li>2. 把待解决的陌生事物组合在一起, 通过类比变为熟悉的事物。例如: 直接类比, 把待解决事物与性质不同但有某些相似的事物直接比较; 拟人类比, 把待解决事物的要素与创新者自身等同起来, 设身处地地想象如何解决</li> <li>3. 把熟悉的事物变为陌生, 以新的视角去观察处理已熟悉的事物, 摆脱习惯性常规的束缚, 发现新联系, 提出新构思。例如: 抽象类比, 把熟悉事物的关键词义或定义变换为抽象词义或一般陈述, 如金属切削(分离、成型); 幻想类比, 把熟悉事物与在现实中难以存在的幻想中的事物类比, 以探求新观念和新构思</li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 把表面上互不相关的因素联系在一起</li> <li>2. 化相识为不相识(同质分化), 化不相识为相识(异质同化)</li> <li>3. 调动人们的潜意识</li> </ol>
	组合创新法	<ol style="list-style-type: none"> <li>1. 把各方面的技术专家组合在一起</li> <li>2. 分析市场需求</li> <li>3. 将两个以上的技术因素组合, 得到有创新功能的技术产物</li> <li>4. 技术因素包括相对独立的技术原理、技术手段、控制方式、工艺方法、材料、动力源等</li> <li>5. 组合方法有: <ol style="list-style-type: none"> <li>1) 性能组合: 将若干产品的优良性能结合起来, 如铜(铁心)线</li> <li>2) 原理组合: 两种以上技术原理组合成复合技术系统, 如喷气原理与燃气轮机技术结合, 得喷气式发动机</li> <li>3) 功能组合: 将具有不同功能的技术手段或产品组合到一起, 形成多功能的技术系统</li> <li>4) 模块组合: 把产品看成若干通用模块的有机组合。根据需要选择不同模块加以组合, 得到不同的设计方案</li> </ol> </li> </ol>
	特点	<ol style="list-style-type: none"> <li>1. 技术领域相互转移渗透, 形成杂交的边缘学科</li> <li>2. 已开发的成熟技术合理组合, 创造出崭新的技术系统、经济有效, 操作方便</li> <li>3. 形式多样, 应用广泛</li> </ol>

## 2 基于功能的设计思维

### 2.1 功能的分类

#### (1) 基本功能和辅助功能

**基本功能** 产品所具有的用以满足用户某种需求的效能, 也就是产品的用途或使用价值。产品的基本功能根据定义的不同可以有一个或数个。

**辅助功能** 与基本功能并存的、次要的、附带的功能。它可以使产品的功能更加完善, 但没有它, 产品并不失去基本使用价值。

#### (2) 目的功能和手段功能

**目的功能** 任一功能的存在都有一定的目的, 因此任何功能, 不论是基本功能或辅助功能都可视为目的功能。

**手段功能** 目的功能往往又是实现另一目的之手段, 相对另一目的来说, 它又是一种手段功能。

任何功能都具有目的功能与手段功能两种性质, 这就是功能的两重性。

#### (3) 使用功能和表现功能

**使用功能** 指产品的实际使用价值。

**表现功能** 在使用功能基础上, 对产品起美化、装饰作用的功能, 它是通过造型设计实现的。

#### (4) 必要功能和不必要功能

**必要功能** 指用户所需要的功能, 包括基本功能和辅助功能。

**不必要功能** 在产品设计中不必要的或过剩的功能, 通常表现在以下几个方面:

- 1) 不适当地加大安全系数;
- 2) 采用的公差、粗糙度超过产品的实际要求;
- 3) 使用了超过功能要求的材料或元件;
- 4) 具有过长的寿命指标;
- 5) 具有不相称的表现功能, 如廉价物品采用豪华的包装、在看不见的部位施以装饰镀层等。

## 2.2 功能分析

### (1) 基本概念

**功能元** 能完成一项或一方面功能的作用单元。它可以是各种层次的功能子系统,并且通常和机械结构中的总成、部件、元器件等概念相对应;也可以是一些最基本的、没有必要再作细分的基本功能单元。

**技术系统** 一个由若干功能元组成的功能系统,通过各功能元之间的逻辑联系和有序作用,实现系统的总体功能。

**功能分析的过程** 将系统的总体功能分解为子系统功能和基本单元功能的过程,这种分解通常从总体功能开始,分层进行,而且每作一步都必须对所建立的功能元作出明确的功能定义,明确其功能的本质,限定其功能的内容,使其在概念上与其他功能相区别。

### (2) 黑箱

在没有获得具体功能解之前,系统用一个抽象的黑箱来表征。能量流、物料流和信息流由黑箱的一侧边界输入,由另一侧边界输出。黑箱虽代表一个还不知道具体解的技术系统,但通过对输入、输出之间的变化,对系统的功能可以作出准确的描述和定义。在系统的功能为定常过程时,只需对输入、输出作量的描述。如果是非定常过程,则必须同时对输入、输出与时间的关系作出描述,系统的功能方能最终确定。功能系统设计的任务就是首先作出这张黑箱图,然后求取箱内的具体解。

作为输入的能量流,可以是机械能、热能、电能、化学能、光能、核能等能量流,也可以是它的某一具体分量,如力流、扭矩流、电流等。例如,作为一个技术系统,内燃机的功能是将输入燃料的化学能流转换为机械能流,核电站的功能是将核能流转换为电能流。

作为输入的物料流,可以是气体、液体或各种形式的固体,如毛坯、原材料、半成品零件、部件、成品等,也可能是粉尘、磨屑、泥沙之类待处理物。机械技术系统将对它们实行结合、分离、移动、气相转液相之类相的变化,以及各相之间的混合等转化。如机床实现多余物料从工件上分离,输送带实现物料的搬运等。

作为输入的信息流,可以是各种测量值、输入指令、数据、图象等;信息的载体可以是机械量、电量、化学量等;信息的形式可以是模拟量,也可以是数字量。技术系统将对信息流作加工、处理或转换。如凸轮机构具有将转动量信息转化为直线运动量信息的功能,计算机能做数据处理,A/D转换器能将模拟量转换为数字量信息,数字式电视机能将接收到的数字信息转化为图象信息。

任何技术系统都可以视为这3种流的处理系统。

有时三流并存,如现代的加工中心和自动生产线,有时只存在能量流与信息流,如电视机。系统在完成功能的过程中,外界将向它输入各种制约和干扰,同时系统也将对外界输出干扰和影响,如振动、噪声、化学排放、热量、废弃物等。

技术系统对能量流、物质流和信息流的处理不仅有性质与数量的要求,而且有质的要求。在输入、输出的描述中同样必须加以明确,如输入系统的电压规定为 $220\text{V}\pm 5\text{V}$ ,输出功率为 $10\text{kW}\pm 10\text{W}$ 。通常对制约与干扰也有规定,如系统的工作温度为 $20\text{C}\pm 3\text{C}$ 。

### (3) 功能定义

1) 功能定义的含义 功能定义是:以简洁准确的语言对整个产品的功能及产品各构成要素的功能一一描述清楚,借以搞清产品及其各个部分所起的作用,以掌握分析对象的本质。

#### 2) 功能定义的作用

① 明确功能构成 产品及其组成部分通常具有多项功能,功能定义可以明确分析对象的功能构成及数量,并准确地加以描述。

② 为功能整理和功能评价建立基础 功能的分类、整理与评价,都是以定义了的功能为对象,而不是以产品及其零部件为对象。

③ 便于构思新方案 将功能从具体的产品中抽象出来,针对价值低的功能实现方案加以突破、创新,以提高产品的价值。

#### 3) 功能定义的要求

① 用准确、抽象和便于定量的词汇 功能定义得准确,有利于给后续工作指明正确方向。用抽象化词汇表达有利于打破旧框框,扩大思路,便于创新。尽量使用可定量名词则有利于对新方案的定量评价。

② 给每一项功能下一个定义 对对象物的每一项功能分别下定义。无遗漏地给每项功能下定义就能逐步看清设计中存在的问题。

4) 功能定义的方法 用一个动词和一个名词定义一项功能,动词作为谓语用以表示一项活动,名词作为宾语表示活动的对象。动词和名词所表达的功能是从如图36.1-1所示的,作为主语的对象物中抽象出来的。这样做的好处是有利于脱离现有方案的框框而创造出新的对象物,用更低的寿命周期成本实现已经定义的功能。表36.1-4是三个对象物功能定义的举例。

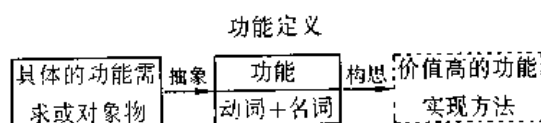


图 36.1-1 功能定义



表 36.1-4 功能定义举例

对象物	功能定义	
	动词	名词
传动轴	传递	转矩
油箱	盛	油
台虎钳	夹紧	工件

5) 重视与功能相关的条件 功能的实现离不开一些具体条件。但是在功能定义中并没有明确描述这些条件,而在对功能具体考察时却不可忽略它们。这些条件可在回答表 36.1-5 中的问题时得到进一步明确。

表 36.1-5 帮助明确功能实现条件的问题

序号	问 题
1	在何时、何处以及什么情况下实现此功能?
2	什么对象发挥此功能?
3	实现此功能的目的是什么?
4	以什么手段实现此功能?
5	要求此功能达到什么程度?

### 2.3 功能综合与整理

把已经定义了的全部功能,按照它们之间的逻辑关系编制出功能系统图以便进一步查明必要功能,发现和剔除不必要功能,找出有待改善价值的功能范围和功能层次,这项工作称为功能综合整理。

#### (1) 功能的逻辑关系— 功能系统图

产品中的任何一项功能都应当有它的目的,同时为实现这项功能可能还需要一项或多项功能作为手段。前者称为后者的目的功能(上位功能),后者称为前者的手段功能(下位功能)。功能之间的这种目的-手段关系称为上、下位关系,为实现同一个目的的功能而作为手段的各个功能之间则为并列关系。

图 36.1-2 表明了台虎钳的某几项功能之间的逻辑关系。其中夹持工件与移动夹头,以及夹持工件与形成夹持力都是上、下位关系。移动夹头与形成夹持力则是并列关系。

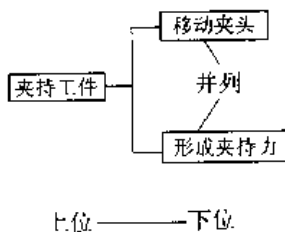


图 36.1-2 功能的上、下位关系及并列关系

按照功能的上下位关系及并列关系,将产品的全部功能系统化,可以建立起产品功能的逻辑关系的体

系。这种体系可绘成如图 36.1-3 所示的树形结构图,称为功能系统图。由图可见,功能  $F_1$ 、 $F_{12}$  的目的是实现功能  $F_1$ , 功能  $F_{21}$ 、 $F_{22}$  以及  $F_{23}$  的目的是实现功能  $F_2$ , 而功能  $F_1$ 、 $F_2$  的目的则是为了实现创新设计所要求的最终功能  $F_0$ 。

功能系统图中的某些功能可以划分为一群相对独立的、群内各功能相互关系密切的功能子系统。图 36.1-3 中的功能  $F_1$  及其下位功能  $F_{11}$  和  $F_{12}$  构成一个功能子系统,而功能  $F_2$  及其下位功能  $F_{21}$ 、 $F_{22}$  以及  $F_{23}$  又构成另一个功能子系统。每个功能子系统即为一个功能领域。

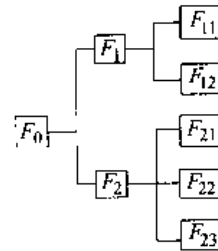


图 36.1-3 功能系统图的一般形式

#### (2) 功能整理的作用

1) 确认必要功能的系统 经过功能整理建立起必要功能的功能系统图,它将是方案设计的基础。

2) 发现不必要功能 产品中与必要功能系统没有联系的功能是不必要功能,消除这些不必要功能是功能整理的重要目的。

3) 划分功能领域 对经功能整理所确定的各功能领域进行分析,可以确立应当在怎样的功能范围和功能层次上进行产品设计。

#### (3) 功能综合整理方法

功能综合整理的有效方法称为功能分析系统技术,其实施步骤如下。

1) 填写功能卡片 将已经定义了的所有功能中的每项功能填写一张功能卡片。功能卡片的内容包括:功能编号、功能定义、对象物名称及成本,见图 36.1-4。卡片放置在能够看到并容易拿取的地方。

编号	对象物名称	成本
功能定义		

图 36.1-4 功能卡片

2) 按目的-手段的关系排列功能卡片 当功能卡片数量多时,把基本功能和辅助功能的卡片分开。首先排列基本功能的卡片,然后排列辅助功能的卡片。通过

回答“为什么需要这个功能?”找出这个功能的目的功能。如果目的功能不能明确,要检查功能定义是否有遗漏或表达不当,必要时补充功能卡片或修正功能卡片。通过回答“怎样实现这个功能?”寻找这个功能的手段功能。上述两个提问代表了上、下两个不同的寻找方向,交替使用两个提问,以检查已经排列出的功能上、下位关系的正确性。

3) 绘制功能系统图 根据功能卡片排列所显示的功能逻辑关系,将上位功能画在左侧,下位功能画在右侧,上、下位功能之间用线段连接起来即构成功能系统图。

绘制功能系统图从连接基本功能开始,首先绘出功能系统图的主要骨架,然后将辅助功能逐步连接上去,以完成功能系统图。

考察功能的目的、手段关系,可能出现一些与功能系统图无直接联系的功能。若某项功能的上位功能不明确,一种原因可能是此项功能定义不当,应予纠正;另一种原因可能是为满足基本功能实现的背景条件而必要功能的上位功能不够明确,这可通过反复考察找出它作为手段功能的适当位置;那些目的始终不能明确的功能是不必要功能,应予剔除。

## 2.4 功能评价

对已经明确的功能领域,查明实现此功能领域所可能做到的(不论在何处)最低成本,并与实现此功能领域的当前实际成本相比较,以此衡量该功能领域的价值,此即为功能评价。通过功能评价找出价值低的功能领域作为重点改进对象。功能评价活动包括:确定功能现实成本,确定功能目标成本,功能评价以及选择改进领域。

### (1) 确定功能现实成本

功能现实成本是指从用户的立场看,以现行方案获得此项功能所需要的成本。通常成本是以产品或产品要素为对象核算的,在功能评价中,则要求以功能为对象计算成本。

如果产品及其要素的成本是已知的,需要以此为依据计算各功能的成本。产品中一个要素往往对实现多项功能有作用,而一项功能也常常需要多个要素协同工作才能实现。可以按照表 36.1-6 中的原则,通过产品要素的成本,推算功能的成本。

表 36.1-6 根据产品要素成本推算功能成本的原则

序号	原 则
1	一个要素只对一项功能有贡献时,该要素的成本全部计入相应功能的成本中

(续)

序号	原 则
2	一个要素对多项功能的实现有贡献时,需要估计该要素对实现各功能贡献大小的比例,将要素成本按此比例分配给相应的功能
3	一项功能需要多个要素协同作用才能实现时,此功能的成本是相应各要素分配给它的成本之和
4	一个功能领域的成本等于所有它所属下位功能成本之和

例 某产品由四个要素组成,成本依次为 300 元、500 元、90 元和 70 元。该产品有三项功能。要素 1 对三项功能都有贡献,此贡献大小所占的比例依次为 60%、10%和 30%;要素 2 对功能 2 及功能 3 有贡献,其比例为 60%和 40%;要素 3 及要素 4 分别只对功能 1 及功能 2 有贡献。则各要素成本可按表 36.1-7 所示的方法分配给各项功能。

表 36.1-7 根据产品要素成本计算功能成本

要素	要素成本/元	分配值		
		功能 1	功能 2	功能 3
要素 1	300	180 元 60%	30 元 10%	90 元 30%
要素 2	500		300 元 60%	200 元 40%
要素 3	90	90 元 100%		
要素 4	70		70 元 100%	
功能成本/元		270	400	290

将各项功能分到的成本求和,所得各项功能的成本依次为 270 元、400 元和 290 元。

根据现实产品各要素成本,推算出的功能成本是功能的现实成本。

### (2) 确定功能目标成本

用货币作为功能的统一衡量尺度,从用户的立场上回答“这项功能值多少钱?”就要考察实现此项功能的各种方法,其中具有最低成本的方法所需要的费用是此项功能的评价标准值,也称为功能必须成本或功能目标成本。以功能目标成本与功能现实成本相比较,可以明确当前实现此项功能方法的价值系数。

确定功能目标成本常用以下方法。

1) 实际调查法 广泛收集实现同样类型功能的已有产品的信息,将实现条件相同或相近的功能,分别以功能达到的程度(功能水平)和功能现实成本为横、纵坐标,绘出如图 36.1-5 所示的散点图。按各功能水

平下成本最低的点画一条回归直线，即为此项功能的最低成本线。以最低成本线上点所对应的成本作为相应水平下功能的目标成本。用这种方法得出的结果实用、可靠。然而即使是成本最低的产品也可能存在改进的余地，据此而确定的目标成本有可能偏于保守，必要时可予以调整。

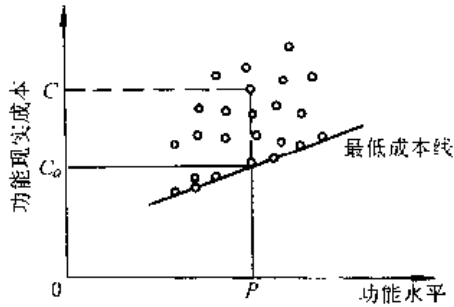


图 36.1-5 功能的最低成本线

2) 经验估计法 依靠专家的知识、经验、创造力和直觉判断力，对功能目标成本做出估计。通常由专家针对已明确的功能构思出几个可行方案，而后对这些方案的成本进行估算，从中找出成本最低的作为功能目标成本。邀请多名专家，按照系统化的咨询方法实施这项工作，可以提高估计的准确性。

3) 理论计算法 在一定条件下，有些功能的实现方法已被明确限定。在这种情况下，功能的目标成本可用相应的工程技术设计公式计算得出。例如：在规定距离间，用机械方法传递一定转矩的成本；经规定路线传送一定电流需要的最低成本都可按工程技术标准的规定进行计算。

(3) 功能评价方法

对功能的评价是定量地确定以功能与功能的寿命周期成本之比表示的功能价值。实际应用中，常以下式的计算结果作为对功能的定量评价。

$$V_c = \frac{C_c}{C}$$

式中  $V_c$ ——功能价值系数；  
 $C_a$ ——功能目标成本；  
 $C$ ——功能现实成本。

价值系数  $V_c$  是一个无量纲的数，其最高值取 1.0，表示功能的价值已达到最高，功能的现实成本已低到此项功能的目标成本，从经济上看，此项功能的实现方法当前已没有改进的余地；价值系数  $V_c < 1.0$  时，表明功能的价值偏低，当前还有通过改进而降低成本的余地。价值系数越小，说明当前的功能实现方法越需要改进。

衡量一项功能的成本可能的最大降低幅度，可使用由下式所定义的指标。

$$C_N = C - C_a$$

式中  $C_N$ ——功能成本可能的最大降幅；

$C$ ——功能现实成本；

$C_a$ ——功能目标成本。

使用功能分析方法实现功能定量评价的过程如图 36.1-6 所示。功能评价中使用的两个指标价值系数  $V_c$  和成本可能的最大降幅  $C_N$  都各有使用的局限性。 $V_c$  是一个相对指标，表明现行的功能实现方法的相对可改善性， $V_c$  值越低，表明现行方案与最低成本方案的相对差距越大，一般也越容易取得改进成果，但它不能反映改进后成本降低额的绝对量。与此相反， $C_N$  是一个绝对量指标，它表明改进成功后，可降低成本的绝对量，但是它不能反映现行方案与最低成本方案的相对差距，因而也不能说明现行方案改进余地的大小。

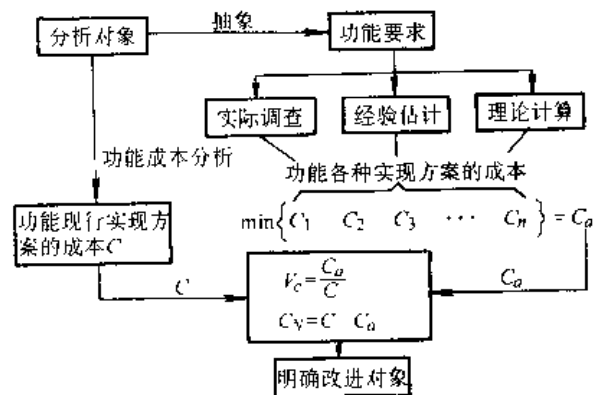


图 36.1-6 功能分析与评价过程

例 某产品有  $F_1$ 、 $F_2$ 、 $F_3$  和  $F_4$  共四个功能领域。各功能领域的现实成本、目标成本以及计算得出的价值系数和可能最大成本降低额，如表 36.1-8 所示。由计算结果知功能领域  $F_2$  的  $V_c = 1, C_N = 0$ ，说明该功能领域的实现方案是合理的，当前无必要更改。 $F_3$  的价值系数最低，说明该功能领域的实现方案问题最明显，但是对这个功能领域做改进，即使得到最好的结果也只能降低成本 23 元，而改进  $F_1$  可能降低 60 元，改进  $F_4$  可能降低 54 元。对比  $F_1$ 、 $F_4$  的价值系数，前者 0.9 后者 0.8，表明对  $F_4$  实施改进可能更易于取得成效。

表 36.1-8 功能评价指标的应用

功能领域	功能现实成本 C /元	功能目标成本 C <sub>a</sub> /元	功能价值系数 V <sub>c</sub>	功能成本可能的最大降幅 C <sub>N</sub> /元
F <sub>1</sub>	600	540	0.9	60
F <sub>2</sub>	88	88	1.0	0
F <sub>3</sub>	50	27	0.54	23
F <sub>4</sub>	270	216	0.80	54

(4) 选择改进领域

经过对功能的定义、整理与评价，已经剔除不必要

功能,并从定性和定量方面查明了价值系数低的功能,并初步考查了功能多种可能实现的方案。在根据功能评价结果选择改进领域时,功能价值系数 $V_i$ 和可能的最大成本降低额 $C_i$ 这两个指标需结合使用,既考虑改进取得成果的可能性大,又考虑使改进成功后的效益高。

选择改进领域不能仅依据功能评价的结果,即使功能价值系数 $V_i=1$ 的功能方案也有可能功能水平及有关条件的满足程度等方面存在许多问题。对于存在问题突出的功能,要从是否能可靠地实现用户所需要的功能这个角度考察并选择改进领域。如果在功能这个本质的方面存在问题,单纯降低成本就没有意义了。

科学技术的发展将使新的功能以及功能实现方法不断出现,功能评价应以动态的观点考察问题,在方案优化过程中,集中力量优先选择那些力所能及的、把握性大的、预计效益高和短期内见效的功能领域作为优化对象。

### 3 基于产品寿命周期的设计思维

#### 3.1 产品的市场寿命周期

一个产品的市场寿命周期分为4个阶段,即:投入期、成长期、成熟期和退让期,如图36.1-7所示。

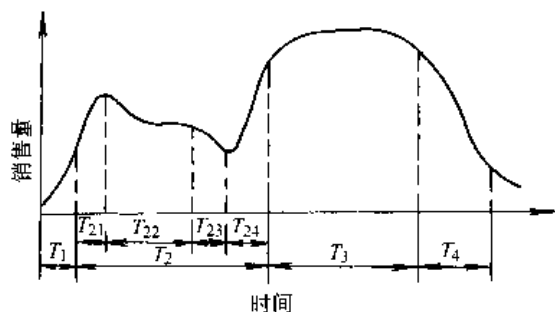


图 36.1-7 产品的市场寿命周期

1) 投入期( $T_1$ ) 产品初次投放市场的阶段。需要用各种手段向用户介绍产品,如展销、参加博览会、作各种媒体形式的广告等,促销费用很大,销售量很低,一般很少取得效益。

2) 成长期( $T_2$ ) 可分为成长前期和成长后期。成长前期又分为试销期、初步稳定期和失望期。

试销期( $T_{21}$ ) 产品开始批量生产和投放市场,由于利用各种方式广泛宣传新产品的优点,用户受到吸引,新产品迅速被市场接受。销售量迅速增长,产品逐渐扭亏为盈。

初步稳定期( $T_{22}$ ) 由于销售量迅速增加,使用者越来越多,经过一段使用过程,新产品的缺陷开始暴露,用户开始抱怨新产品有缺点。虽有不利于新产品的

议论,但肯定性意见是基本的,后期销售量会略有减少,但对企业的盈利影响不大。

失望期( $T_{23}$ ) 新产品的缺陷全面暴露,用户意见纷纷,有些用户可能对新产品持否定态度,有些保守的用户可能仍去购买功能相同的老产品,有些原来想买新产品的用户可能改持观望态度。这时形成购买新产品的低潮,导致市场需求量急剧下降,企业盈利受到较大影响。

成长后期( $T_{24}$ ) 生产厂在听取用户意见的基础上,针对新产品存在的缺陷,大力进行改进,使产品的优点更为突出,缺陷得到弥补。通过宣传试用,广大用户对新产品又有了较好的评价和信任,便进入了成长后期。产品的销售量迅速回升,企业盈利额迅速增加,逐步达到高峰。

3) 成熟期( $T_3$ ) 产品在市场上的供求基本平衡,销售量趋于稳定,这个阶段时间较长,可一直持续到下一代新产品的出现。

4) 退让期( $T_4$ ) 由于新的替代产品的出现,原产品便进入了衰退时期。在这个阶段,或者由于习惯的作用,保守的用户不会很快接受新产品,或者由于新产品投入市场的数量有限,供不应求,所以老产品不可能一下子就没有销路,要经过一个新老产品共存,新产品销售量逐步上升,老产品销售量逐步下降的退让阶段,老产品才会完全退出市场。

产品价格、销售额、利润在整个市场寿命周期内的变化情况如图36.1-8所示。根据产品所处市场寿命周期阶段的不同,进行分析思考的重点也不同。

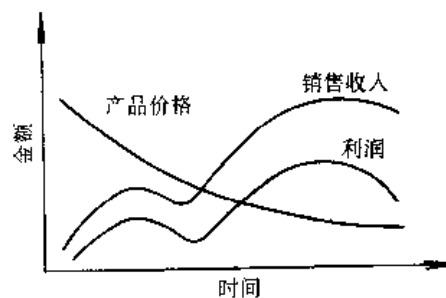


图 36.1-8 产品价格、销售额、利润在市场寿命周期内随时间过程的变化

1) 当产品处于成长期的前3个阶段时,重点问题是根据用户的意见,改善产品的功能,提高产品的质量,以加强产品的市场地位;

2) 当产品处于成长后期和成熟期时,重点应放在成本降低上,以提高产品的经济效益;

3) 当产品处于成熟阶段末期或刚刚进入退让阶段时,重点要尽可能地延长产品的市场寿命。这时,可通过改变产品的外观,局部提高产品的某些功能,或使产品成本有大幅度的降低,以维持产品的市场占有率,

推迟产品的衰退时间。

### 3.2 产品的全寿命周期

产品的全寿命周期是指从制造该产品的原材料开采制备起,到产品使用寿命终止的全部过程。现代设计要求在产品设计阶段必须对这一全过程中出现的问题作出全面的、综合的考虑。图 36.1-9 是产品的全寿命周期阶段构成。

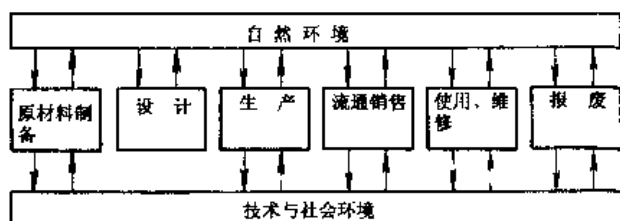


图 36.1-9 产品全寿命周期的构成

1) 原材料的开采与制备阶段 从经济性、材料本身性能和注重材料资源保护与再生以及材料开采、制备过程中污染、毒性和能源消耗等方面选材。材料的选择与很多社会因素相关,在设计中都应予以注意。

2) 产品设计阶段 产品设计的水平直接影响产品的功能、成本、质量、使用和研制周期。统计资料表明,产品的质量事故中有 50% 是由于设计失误而造成的,产品的成本 60%~70% 取决于设计。

3) 产品生产阶段 产品应具有良好的工艺性能,特别是应该和当今的先进生产设备和技术,包括生产组织与管理技术相适应。在生产过程中要尽量少消耗能源,减少对环境的污染,所需要的劳动条件要符合社会劳动法规的要求。在设计中,对配件从国际市场的获取、国际市场上的规格、标准、为便于异地生产而将产品作必要的模块分隔等,都应作出周密考虑。

4) 产品销售阶段 产品的流通、销售在设计中应予以重视,特别是运输过程中的保护;为便于运输应将产品设计成便于拆装的形式;产品应设计成易包装的结构;对包装材料的污染问题应引起社会的特别重视。

5) 产品的使用阶段 一个好的产品设计,要求具有高效率、很好的可靠性、第一流的售后服务。

6) 产品的报废与消亡阶段 产品使用寿命结束后,“遗体”将转化为环境的负担。要求产品在设计中就对报废产品的回收与再利用作出周密考虑,形成现代设计中的一个新的分支——面向拆卸与回收的设计方法。

## 4 基于过程的设计思维

### 4.1 设计任务的类型和进程

#### (1) 设计任务的类型

设计任务一般可分为如下 4 类:

1) 自主开发设计 企业或个人,根据市场的需要,或根据突发的创新灵感和对未来应用价值的预测,在原理方案未知的前提下,所作出的创造性的、全新概念的设计。其成果通常可列入发明之列,并取得专利。开展这项工作同样会有一个计划,但不像合同与订单那样在时间与交货期方面存在严格的制约,这对充分展开创新思维是非常有利和必要的。

2) 接受订货开发设计 这是根据用户订货要求所进行的开发设计。它常常是满足用户个别需要的专用非标设备。设计部门必须承担一定的风险,所以必须进行慎重的论证,主要技术应在自己熟悉的业务领域内,大多数技术和所用零、部件都应是成熟的,设计与制造周期、交货时间都应与自身力量适应。对使用尚不熟悉的新技术要作充分的可行性论证,而且不宜太多。用户通常采用招标方式寻求制造商,能否中标则取决于投标方的综合实力。

3) 适应性设计 在总的方案和原理解不变的前提下,为适应用户某些新的要求,或克服产品市场调查所反映的缺点,所进行的发展性设计。它常常对个别功能元或零部件作出增补或改进。

4) 变型设计 是一种仅作尺寸、布置方式、能力大小等变更,基本上不牵动系统功能结构的设计。如对齿轮减速箱作系列设计,发动机作四缸、六缸、直列、V 型等改型设计等。

#### (2) 设计任务的进程

图 36.1-10 为一张普遍适用的设计工作流程图,它将设计划分为 4 个阶段:

1) 制订任务书阶段 编制出详细的要求表,对所设计系统的功能、制约条件和重要性提出明确的要求。它在随后的各个工作步骤中可以补充或调整,使它始终保持最新最合理状态。但这种改动在企业内部,在企业与订货单位之间,都应有严格的程序。

2) 方案设计阶段 根据任务书,通过抽象化将设计任务提到本质高度加以认识;通过功能分析与综合,建立功能结构;通过寻求合适的作用原理将功能结构转化成具体的作用结构;给出系统的原理解。通过原理方案评价中选的方案,在随后的设计中还必须继续对其“优化性”作跟踪考评。

3) 详细设计阶段 为所设计的系统在结构方面确定一个解,包括选用材料,确定形状和尺寸,确定具体参数等。如果在方案设计阶段产生几种原理解,则必须同时完成几个技术设计方案,然后在更具体、更高的信息阶面上对方案作出技术经济评价,通过相应地组合或移植这些局部解,构成最后的结构解。它的表现形式通常是详细的总图和部件图,以及完整的计算资料。

4) 施工设计阶段 形成最终的制造信息。它可能是一套零件图,对每一零件提供了详细的形状、尺寸、公差、表面质量、材料与热处理、表面处理等信息,也可能是一块印刷线路版的设计图样。

5) 优化过程 设计过程中的原理设计的优化、结构设计的优化和制造的优化。它们通常贯穿 2~3 个设计阶段,并且通过反复的循环工作向最佳方案逼近,而

价值系数自始至终是评价的准星。

(3) 批量产品的开发进程 (见图 36.1-11)

(4) 不同的设计类型在设计进程的各阶段的工作重点

1) 自主开发类设计的阶段重点 必须抓好市场预测和需求分析,形成企业的中、长期开发计划和近期开发目标。特别要抓好情报工作,才能形成很好的设计

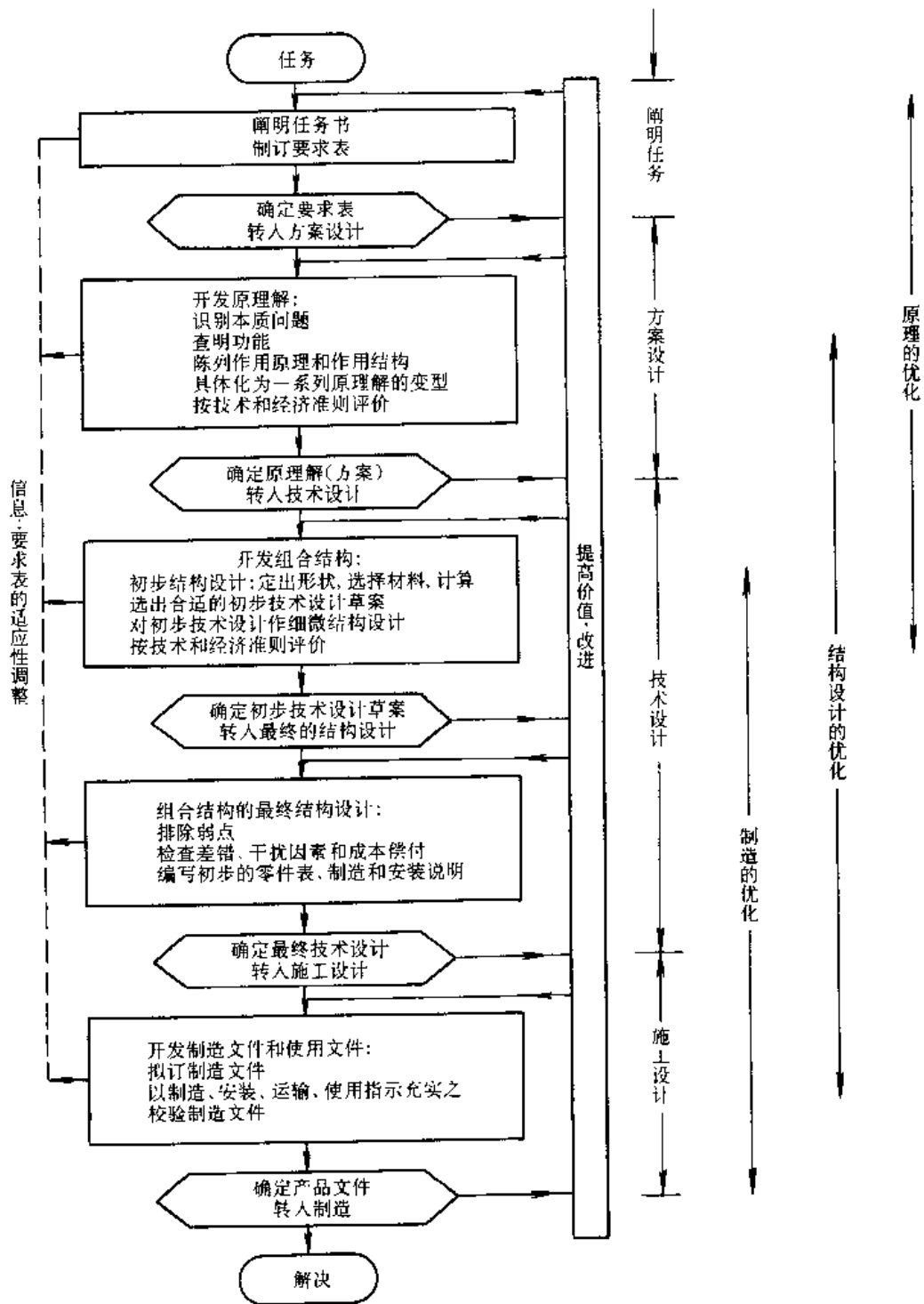


图 36.1-10 设计工作流程图

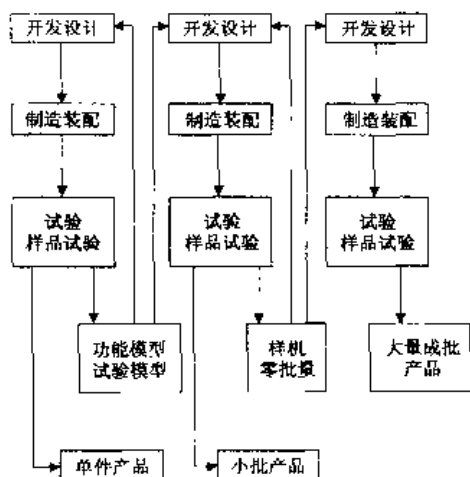


图 36.1-11 批量产品的分步式开发进程

任务书。任务书确定之后应集中全力求取系统的原理解，它通常取决于企业的科技实力。在现代企业中，大都建有 R&D (Research and Design) 部门，原因正在于储备技术。技术设计和施工设计借助 CAD/CAM 可以高效而准确地完成。设计最终的成功与否主要取决于市场反映。

2) 接受订单开发设计的阶段重点 情报工作的重点是捕捉订单，详细分析招标单位的任务内容，快速形成自身的基本计划与标书。由于设计任务属于企业擅长技术的引伸，因此在企业的资料库中应有大量的、各种层次功能解的方案，原理设计应以此为基础，用较快的速度完成，甚至很多零、部件现有结构设计可以移植到新设计中。设计结果则通过产品交货和用户验收效果最后定论。

3) 适应性设计的阶段重点 情报工作的重点应是原产品的市场反映。通过分析这些反映形成产品改进设计的任务书，它通常是对系统的局部功能作增补或改进，这类设计各阶段的工作都应集中在改动部分。

4) 变型设计的阶段重点 情报工作的重点是市场对变型的需求信息，在综合这些信息的基础上作出变型设计决策。往后的设计工作将主要进行结构与工艺设计，但不排除对系统的功能结构作局部变动。

## 4.2 设计任务的形成与决策

### (1) 形成设计任务的外因

1) 消费者的潜在需要 各种消费阶层，各种消费群体都会有潜在的需要，挖掘、发现这种需要，创造一种产品予以满足，是一切创新设计的出发点。

2) 用户对现有产品的反映 现有产品的市场反映，特别是用户的批评和期望，是企业必须关注和应迅速作出改进的重点。

3) 产品市场寿命周期产生的阶段要求 当已有产品进入市场寿命周期的不同阶段后，产品必须不断进行自我调整，以适应市场。

4) 竞争对手的技术挑战 市场上竞争对手的产品状态和水平是企业情报工作的重心。

5) 技术发展的推动 新技术、新材料、新工艺对市场上新产品具有很大的冲击。

6) 社会的需要 市场是社会的组成部分，很多政治、军事和社会学问题都通过市场，对产品提出需求。

### (2) 调查和预测

一般企业都有自己的短期、中期和长期的产品发展预测，表 36.1-9 是不同预测对象的预测时期概念。

表 36.1-9 不同预测对象的预测区间

预测对象	预测区间		
	短期预测	中期预测	长期预测
产品预测销售	1~6 个月	6 个月~1 年	1 年以上
经营预测	1~3 年	3~8 年	8 年以上
技术预测	1~5 年	5~15 年	15~50 年
国民经济预测	1~5 年	5~15 年	15 年以上
环境预测	1~10 年	10~20 年	20 年以上

调查是作出预测的主要手段，调查的方法很多，这里列举如下方法：

1) 用户直接需要调查法 这种方法有 3 个要点，即：调查对象范围，调查方法和调查表的设计。

① 调查对象范围 又称抽样范围。当产品使用对象没有明显类别时，可采用简单的随机抽样法，任意从一批对象中抽样调查。若产品的使用对象具有类别差异时，应采用分层比例抽样法，将抽样总数按类作比例划分，然后按划分数字在各类对象中作随机抽样调查，最后作汇总统计。有时可采用整群抽样法，即将对象按群体划分，如按行政区分群，按工业部门分群，按行业分群等，然后从中抽样调查。

② 调查方法 信访、走访、借助会议、甚至街头拦访等，最好对被访者给予相应的奖励。

③ 调查表的设计 表中问题应简明扼要，最好列出可能的答案让被调查者通过打√予以回答，不要使被调查者感到厌烦。

### 2) 德尔非法

① 把预测内容编成一份调查表，表中所列问题要明确、清楚、单一、便于回答，有统一的评估方

法。

② 选定10~50位专家，视情况可增加，这些专家对预测问题有专门的科学技术知识和丰富的实践经验。专家的权威性关系到预测是否能取得成功，需要谨慎选择。

③ 将调查表邮寄给各位专家，进行背对背征询意见，对专家姓名严格保密，以消除各专家间相互影响，便于专家独立思考对问题作出评价，然后将各专家意见及时回收，进行定量统计归纳；再将归纳的意见反馈给各专家，让专家们了解全体意见，在此基础上请专家重新作出判断。第二轮调查表回收后再作统计归纳，然后再次将归纳的意见发给专家，开始第三轮调查。经过多次反复，专家们的意见逐渐趋向基本一致，最终得到满意结论。但也有可能专家们的意见在一定的分散程度上僵持不变，它也客观地反映了专家们的意见分歧。

还可以用趋势外推法进行预测，如根据已知的数据，利用现成的数学拟合方法构造出相应的曲线或关系式，然后外推出所需的预测数据。这方面商品化软件很多，如SPSS等。

(3) 可行性分析的内容与方法

1) 可行性分析的内容

- ① 根据市场分析，确认该开发项目的必要性；
- ② 如市场上已有同类产品，则应详陈其国内外

现状及水平，论证自己新开发产品的先进性；

- ③ 确定产品的技术规格、性能参数和约束条件；
- ④ 归纳、提出新产品开发的技术关键，解决途径，以及本企业相应的资源能力，作出比较论证；
- ⑤ 预期达到的技术、经济和社会效益，预算投资费用及回报率；
- ⑥ 项目进度和期限。

2) 可行性分析的方法

① 产品-市场矩阵 如图36.1-12所示，它详细列出新的市场与新开发的产品，然后思考是用老产品开发新市场（区间Ⅰ），还是开发新产品占领老市场（区间Ⅱ），还是开发新产品占领新市场（区间Ⅳ），最后一种是开发力度和风险都最大的区间，但成功后效益也往往最大。

② 需求-优势矩阵 图36.1-13为一需求-优势矩阵。它由强至弱依次将企业的优势和市场需求分别列于表的行与列标题栏中，然后根据优势项是否与需求项相关在相应格内作记号。显然记号密集的左上角区是新产品优先考虑的范围。

(4) 新产品要求表

一旦对新产品开发作出决策，就应对新产品作出明确的定义，列出对新产品的要求表，向设计部门下达内容明确详尽的设计任务书。表36.1-10为要求表的典型实例。

		当前的市场				新的市场				
		火力发电厂	电动机工业	设备工业	...	水力发电厂	收音机	电视机	家用设备	...
当前的产品	电子计数器									
	测热计		I						II	
	测量计									
	⋮									
新的产品	计时器									
	测压计			III						
	测温计								IV	
	机械功率									
	⋮									

图 36.1-12 产品-市场矩阵

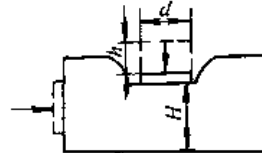


		企业的优势和能力, 按渐降次序排列				
		电子测量	电功的机械测量	光电技术	分层技术	结合技术
顾客市场的需求, 按渐降次序排列	微处理机程序控制	X			X	X
	热量操纵	X	X			
	远距离监控	X		X		
	里程计数	X		X		
	风向和风速测量		X			

图 36.1-13 需求 优势矩阵

表 36.1-10 要求表典型实例

		要 求 表 用于: 油箱贮量测定仪传感器	张: 1 页: 1
改动	必/愿	要 求	负责者
	必	1. 容器、接头、距离 容积: 20~160dm <sup>3</sup> 容器形状任意给定 (形状稳定)	
	必	材料: 钢或塑料	
	愿	容器上的接头: 法兰连接, 形式 A, DIN 75612	
	必	接头在上方	
	必	接头在侧面	
	愿	H=150~600mm	
	愿	d=φ71mm, h=20mm	
	必	容器到显示仪的距离: ≠0m, 要求 3~4m	
	愿	1~20m	
	必	2. 容器内容、温度范围、物料 介质            测量范围    仓库范围 汽油或柴油    -25~65 C    -40~100 C	
	愿	3. 测量信号、能量 传感器输出: 电信号 (贮量变化时电压变化)	



(续)

		要 求 表 用于：油箱贮量测定仪传感器	张：1 页：1
改动	必/愿	要 求	负责者
	必	具备的能源：直流电，电压 6V，12V，24V 电压公差：-15%~25%	
	必 愿	测量公差：输出信号，与最大值之比 ±3% ±2% (连同显示：±5%) 这是指正常行驶时：水平路面， $v$ =常数，具有平常路面所引起的跳动	
	必 愿	灵敏度：最大输出信号的 1% 最大输出信号的 0.5%	
	愿	液面倾斜不影响信号 信号可标定 信号可在容器装满时标定 可测量的最少贮量：最大值的 3%	
	愿	通过特殊信号显示贮存量	
		4. 工作状态	
	必	行驶方向上的加速度：不超出 $\pm 10\text{m/s}^2$	
	必	与行驶方向垂直的加速度：不超出 $10\text{m/s}^2$	
	必	与路面垂直的加速度（跳动）：不超出 $30\text{m/s}^2$	
	愿	行驶方向上的无损冲撞：不超出 $-30\text{m/s}^2$	
	必	行驶方向上的倾斜：不超出 $\pm 30^\circ$	
	必	与行驶方向垂直的倾斜：最大 $45^\circ$	
	必	容器不受压力（接通空气）	
		5. 测试条件	
	必	按照买主指示，对内件和外件作盐雾试验 (注意 DIN 90905)	
	必	容器的测试压力：30kPa	
		6. 传感器的寿命和耐用性	
	必	在受到所测量介质和汗水腐蚀的条件下有 5 年寿命	
	必	在计入汽车全部载荷的情况下，传感器能可靠工作	
		7. 制造	
	愿	传感器尽可能易于改装到各种容量的容器上	
		8. 使用、维护	
	愿	外行人也会装用	
	必	传感器可更换，且无需保养	
		9. 个数	
		可改装的传感器，每天 10000 个 畅销的品种，每天 5000 个	
		10. 成本	
		制造成本 $\leq 3$ 马克/个	
Dhz/Sch73. 6			

### 4.3 方案设计

#### (1) 方案设计的方法学进程

方案设计是为系统确定一个原理解。图 36.1-14 给出了方案设计的一般的方法学进程。

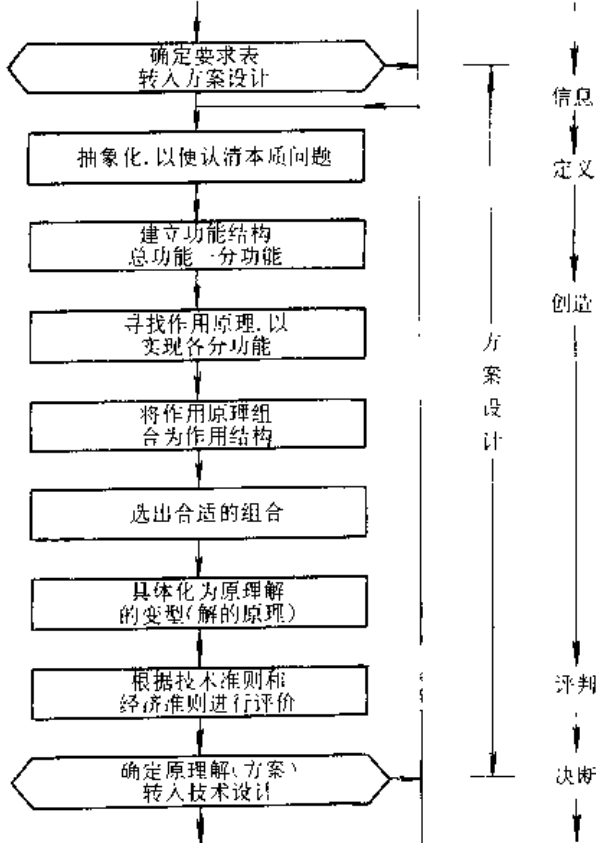


图 36.1-14 方案设计的方法学进程

#### (2) 设计任务的抽象化进程

将设计任务的要求表进行分析，确定所要求的功能和主要条件，撇开与功能和本质无直接关系的约束条件和必达要求，从包含大量定量的任务说明中提炼出定性的任务核心，在不偏向某种解的、抽象的阶面上重新定义设计的根本目标，是方案设计的第一步。抽象化的思维进程可以在分析要求表的基础上一步步通过提炼的方法进行。

如，对于开发汽车油箱贮量测定仪传感器的要求表，就可以像表 36.1-11 那样，抽象化为最后的定义。

#### (3) 功能分解

针对抽象化的功能定义，可以按功能系统设计的要求作出系统的功能结构图。在系统各分支的终端(末端功能)通常为具有独立功能的功能元。它可以分为物理功能元，数学功能元和逻辑功能元三大类。

1) 物理功能元 它反映系统中物理量转化的基本动作。如图 36.1-15。其中最常用的 5 种为：

① “变换—复原”功能元。包括各种类型能量之

表 36.1-11 设计任务的抽象化进程

(按表 36.1-10 所示汽车油箱贮量测定仪传感器的要求表)

第 1 步抽象化结果——消除具体愿望和非本质约束条件

- 体积：20~160dm<sup>3</sup>
- 容器形状任意给定(形状稳定)
- 接头在上方或侧面
- 容器高度：150~160mm
- 容器到显示仪的距离：≠0m，要求 3~4m
- 汽油或柴油，温度范围：25~65℃
- 传感器输出量：任意测量信号
- 外来能量：(直流电 6V, 12V, 24V, 公差±15%~25%)
- 测量公差：输出信号，对最大值之比±3% (连同显示±5%)
- 灵敏度：最大输出信号的 1%
- 信号可标定
- 可测量的最小贮量：最大值的 3%

第 2 步的结果——从定量到定性

- 不同的容积
- 不同的容器形状
- 各种接头方向
- 不同的容器高度(液面高度)
- 容器到显示仪的距离≠0m
- 液量随时间而变
- 任意信号
- (有外来能量)

第 3 步的结果——扩大认识

- 不同的体积
- 不同的容器形状
- 在不同的距离上显示
- 测量液量(随时间而变的)
- (有外来能量)

第 4 步的结果(课题表述成文)——作不偏向某种解的定义

连接测量并显示任意形状容器中不同大小且随时间而变的液量

间的转变、运动形式的转变、材料性质的转变、物态的转变及信号种类的转变等。

② “放大—缩小”功能元。指各种能量、信号向量(“力、速度”等)或物理量的放大及缩小，以及物料性质的缩放，如压敏材料电阻随外压力的变化。

③ “连接—分离”功能元。包括能量、物料、信号同质或不同质数量上的结合。除物料之间的合并、分离外，流体与能量结合成压力流体(泵)的功能也属此范围。

④ “传导—绝缘”功能元。反映能量、物料、信号的位置变化，传导包括单向传导、变向传导，绝缘包括离合器、开关、阀门等。

⑤ “贮存—提取”功能元。体现一定时间范围内保存的功能。如飞轮、弹簧；电池、电容器等，反映能量的贮存；录音带、磁鼓反映声音、信号的贮存。

2) 数学功能元 它反映数学的基本动作，如加和

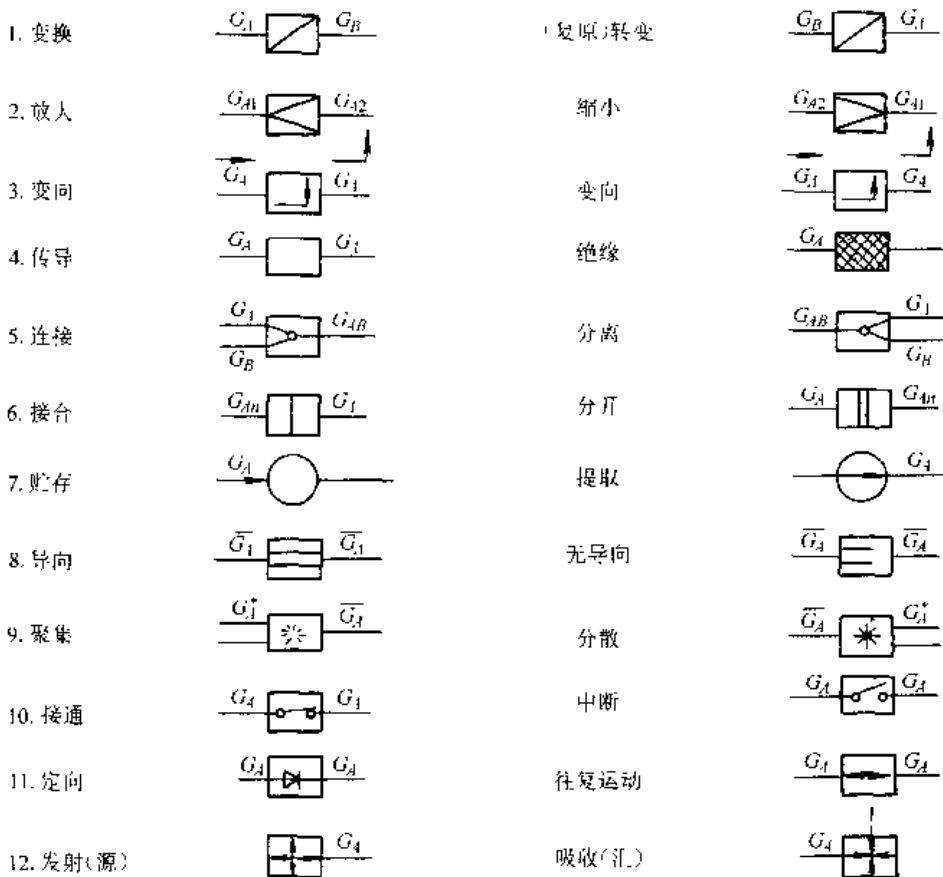


图 36.1-15 物理功能元

减、乘和除、乘方和开方、积分和微分，图 36.1-16 表示数学基本动作。数学功能元主要用于机械式的加减机构和除法机构，如差动轮系；控制系统，如用脉冲计数装置控制动作幅度等等。

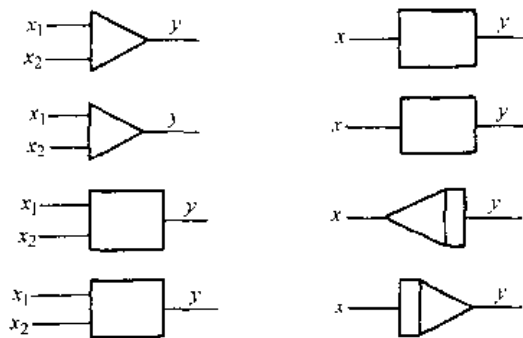


图 36.1-16 数学功能元

3) 逻辑功能元 包括“与”、“或”、“非”元的逻辑动作，主要用于控制功能。基本逻辑关系如图 36.1-17 所示。



图 36.1-17 逻辑功能元

(4) 分功能求解

广泛应用各门学科的原理于特定技术目的，形成

技术原理。按技术原理组织功能载体，在一定条件下作用于加工对象，成为技术分系统，实现分功能。

1) 科学与技术的关系 科学源于技术又是技术的指导；技术以科学为基础又是科学发展的动力。科学成果应该快速转化为技术成果；技术成果应该迅速上升为科学成果。

设计人员的思维要发散，机械设备的设计不应局限于机械学范围，力学、电子学、电磁学、热学、光学、仿生学……都应考虑。

2) 技术原理 应用科学原理构思技术原理的基本途径如下：

- ① 把科学原理转化为技术原理。
- ② 从技术经验中提炼技术原理。
- ③ 由已有技术原理形成新的技术原理。

3) 功能载体 以它的某种特性来完成某功能（物理化学特性、运动特性、几何特性、力学特性）。

功能载体具有易于被认识的显特性，也有待认识的潜特性。发掘功能载体的潜特性往往有助于创新构思。

4) 加工对象的特性 特性随各种条件而变化，要通过试验和分析来彻底掌握加工对象的特性。

5) 工作条件 弄清工作条件是常规工作条件，还

是非常规工作条件。

6) 分功能求解重点

- ① 分析事物内部矛盾, 识别问题的本质。
- ② 寻求实现某功能的技术原理, 加以实验验证。
- ③ 逐步变换工作条件, 适应任务要求。

(5) 设计目录

对于一个功能系统来说, 并非所有功能元的载体都是创新的, 其中大量是存在现成解的。这些解被汇编成解法目录, 是设计中的重要参考。

1) 对象目录 不针对具体任务的一般信息目录, 如材料特性、物体运动特征等。每一对象有分类部分、主体部分、检索部分, 如最多有四个构件的约束运动机构; 分类部分有转动副、连架运动副; 主体部分有运动原理图; 检索部分有连架运动副适于变换运动或增力、构件数、结构耗费等。

2) 作业目录 包括系统性程式设计范围内的各种作业和工作过程、使用条件和判别准则, 如构件变异的作业目录, 一般参数可为几何参数、材料参数; 作业对象可为形状、位置、数目; 具体参数可为面、零件、定性定量分析; 作业内容可为变换、增减、改变; 作业定义可为变异一个面的形状, 变异零件之间的相对位置,

变异一种材料的化学成分等。

3) 解法原理目录 特定功能、特定任务的解法(边界条件)是功能要求、解法原理和功能载体的综合, 如功能要求为固体物料分离, 解法原理中, 分离特征为密度(摩擦因数、导电性等), 利用浮力效应(摩擦效应、库仑效应等)。功能载体是密度介于二种被分离物料密度之间的液体。方法是将物料放入液体中, 轻者上浮, 重者下沉, 达到分离目的。

4) 解法目录 提供各种特定功能的不同解法, 如力的产生解法的解法目录, 分类部分中, 力的类型有重力、惯性力; 有关定律有万有引力定律、牛顿定律; 特性效应有地心引力、离心力。主体部分有力的计算公式、图例。检索部分有力效应的物质条件、产生的场景, 是否需要能量持续输入, 力的作功能力, 力的大小和特征值, 结构参数, 力效应的几何条件, 持续时间和运动条件等。

5) 操作目录 例如, 项目有功能面变换、结合方式变换、形状变化、组合方式变化。变换用各种图形表示, 如形状有平面、曲面、圆柱、四面体、球等。位置变化有外表面和内表面。组成形式变化如接触面之间有滚动、滑动等。

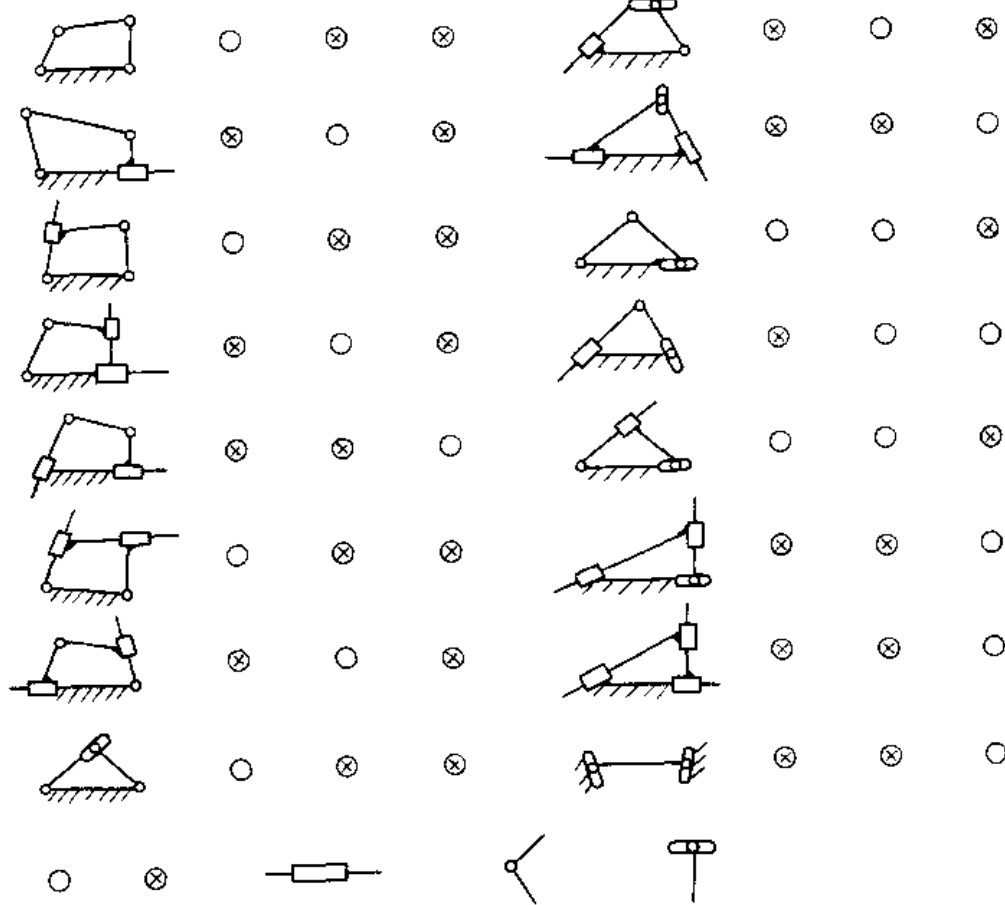


图 36.1 18 四杆机构运动副转换解法目录

6) 功能载体词典 把专业常用功能编为词典。一般关键词为动词,再增加限制性形容词,按抽象程度编排。把技术知识作为素材,归入适当条目之下供参考。启发联想及类比推理使用。

7) 各种手册等。

图 36-1-18~20 举出这种解法目录非常小的一部分实例。

(6) 方案综合

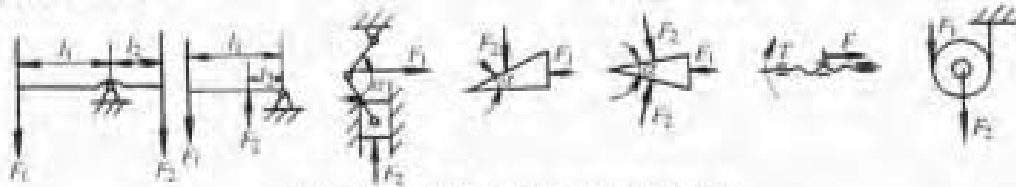


图 36-1-19 机械一次增力功能元解法目录

原理解 功能	机 械				流 气	电 磁
	凸轮传动	联杆传动	齿轮传动	连杆/压解 方式传动		
转 变						
缩 小 (放大)						
变 向						
升 高						
力	静 力					
	动 力					
生	摩 擦 力					

图 36-1-20 部分常用物理基本功能元解法目录

解法 功能元	解法					
	1	2	3	4	...	n
$F_1$	$L_{11}$	$L_{12}$	$L_{13}$	$L_{14}$	...	$L_{1n}$
$F_2$	$L_{21}$	$L_{22}$	$L_{23}$	$L_{24}$	...	$L_{2n}$
...	...	...	...	...	...	...
$F_m$	$L_{m1}$	$L_{m2}$	$L_{m3}$	$L_{m4}$	...	$L_{mn}$

图 36-1-21 原理方案的形态学矩阵

由于总功能最终分解为  $m$  个功能元,而每一个功能元都有可能存在  $n$  个原理解,因此经排列组合就有  $m \times n$  个方案。可以如图 36-1-21 作出一张原理解的“形态学矩阵”。设计者面对该形态学矩阵可以作出第一轮切选,构成几种可行性方案,经评价后,确定最终进入技术设计的原理方案。

4.4 详细设计

图 36-1-22 给出详细设计一般的方法学进程。

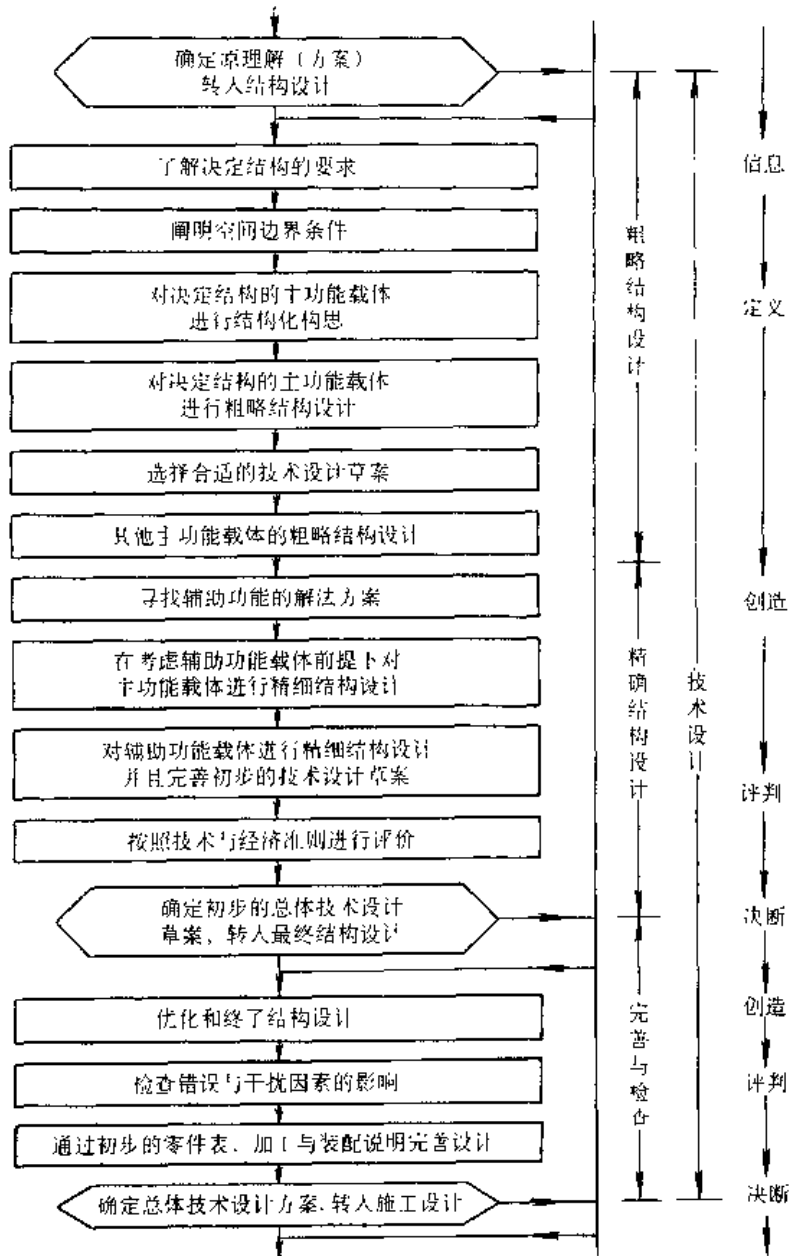


图 36.1-22 详细设计的方法学进程

## 5 产品总体设计的评价与优化

### 5.1 总体设计优化的目标与过程

#### (1) 总体设计优化的目标

提高产品的功能值,降低成本,寻求最大的价值系数,是总体设计优化的目标。

#### (2) 总体设计优化的过程

1) 确定优化对象 通过分析找出关键点,通过对关键指标、关键零部件、关键系统的优化,带动全局的优化。

2) 优化方案的建立 情报是优化方案建立的基础,功能分析是优化方案建立的必经步骤。

3) 优化方案的评选 主要功能处于待完善状态,主要功能是评价重点;主要功能满足的情况下,辅助功能是评价的重点;经济性能永远是评价的重点。

### 5.2 优化对象的确定

表 36.1.12 是选择优化对象时应考虑的方面。

#### (1) 从技术角度选择优化对象

1) 经验分析法 邀请多位熟悉企业生产、工艺和产品设计的专家,通过集体研究,凭借专家个人的经验及掌握的大量信息,对企业的生产过程等各种技术问题和产品的设计情况进行全面的分析并确定优化对象。

优点:简便易行,无需特殊训练,具有相当的准确

性。但定量不足，受到专家个人主观因素影响较大。往往与其他方法联合使用。

表 36-1-12 选择优化对象时所应考虑方面

考虑方面	对象特点
设计方面	— 环节多、结构复杂
	— 设备性能落后、生产技术落后
生产方面	— 原材料贵重、消耗量大
	— 产品产量大
	— 工艺复杂或落后
销售方面	— 燃料、动力耗量大
	— 用户意见大
	— 产品市场竞争激烈
质量方面	— 产品市场趋于饱和
	— 产品返修率高
	— 次品率、废品率高
效果方面	— 产品赔偿率、退换率高
	— 效果显著
	— 具备各种改善条件
难易方面	— 有改善潜力
	— 情报资料齐全
	— 无需大量人力物力财力
	— 牵涉面不广

2) 综合分析法

① 通过对产品进行技术分析，找出能反映产品技术水平的各个指标并确定其在产品中的重要程度；

② 根据产品的具体运行情况对各项技术指标打分；

③ 将产品的各项技术指标得分值与其权重相乘，得出各项指标的加权分值；

④ 将各项技术指标的加权分值汇总，据其高低确定优化对象。

实例分析：

某产品有 A、B、C、D 4 个组成部分。经过企业有关人员分析，决定以可靠性、操作性、维修性、工艺性、生产效率和安全性等 6 项指标来评价每一部分的技术水平，并根据 6 项指标对产品的不同重要性赋予不同的权重，列于表 36.1-13。然后，邀请有关技术专家按 10 分制对各组成部分按上述指标评分，其结果列于表 36.1-14。最后，将各指标得分与权重相乘并累加，得各组成部分的加权总分值，其结果列于表 36.1-15。根据各部分的加权总分值的高低，决定将最低项 A 列为优化对象。

(2) 从经济角度选择优化对象

1) 成本百分比法 这种方法主要是根据产品各部分费用所占总费用的比重，选择比重较大的作为优化

表 36.1-13 各技术指标的权重表

技术指标	可靠性	操作性	维修性	工艺性	生产效率	安全性
权重 (%)	25	10	15	15	30	5

表 36.1-14 各技术指标评分表 (10 分法)

组成部分	技术指标					
	可靠性	操作性	维修性	工艺性	生产效率	安全性
A	5	7	8	7	6	9
B	6	3	9	6	7	8
C	8	10	8	10	8	10
D	8	8	8	9	9	8

表 36.1-15 指标加权评分表

组成部分	技术指标						加权总分
	可靠性	操作性	维修性	工艺性	生产效率	安全性	
A	1.25	0.70	1.20	1.05	1.80	0.45	6.45
B	1.50	0.30	1.35	0.90	2.10	0.40	6.55
C	2.00	1.00	1.20	1.50	2.40	0.50	8.6
D	2.00	0.80	1.20	1.35	2.70	0.40	8.45

对象。这是一种定量方法，较易实行，但由于只考虑了成本一个因素，所以具有较大的片面性。因此在实际运用过程中，一般改进为通过分析 2 个以上的经济指标来选定优化对象。

实例分析：

某企业有 3 个产品项目 A、B、C，它的成本和利润比例列于表 36.1-16。

表 36.1-16 成本利润比例表

项目	A	B	C
成本/万元	16	17	20
成本比例 (%)	30	32	38
利润/万元	15	8	7
利润比例 (%)	50	27	23

由表 36.1-16 可以看出，C 项目的成本占总成本的 38%，而其产生的利润却只占总利润的 23%，因此可以选择 C 项目作为优化对象。

2) ABC 分类法 这是由意大利经济学家帕莱托应用数理统计分析的方法发展而来的。它根据不同零部件(或生产环节)的局部成本所占总成本比重大小来选择优化分析对象。因为每个零部件(或生产环节)的功能不同，其成本分配很不均匀，通常把数量占总数 10%~20%，而成本却占成本总额约 70%~80% 的零部件(或生产环节)称为 A 类项目；把数量占总数 60%~80%，而成本却只占成本总额约 10%~20% 的零部件(或生产环节)称为 C 类项目；其余的称为 B 类项目。ABC 法就是把零部件(或生产环节)按成本比重大小顺序排列，选择 A 类项目作为重点分析对象。



实例分析:

某产品由 10 个零部件组成, 试用 ABC 法进行分类。

解: 把各零部件从 1 到  $N$  ( $N$  为正整数) 编号, 则有

$$\text{第 } i \text{ 个零件的成本比重} = \frac{\text{第 } i \text{ 个零部件的成本}}{\text{所有零部件成本之和}}$$

表 36.1-17 某产品成本 ABC 分类表

零件编号	10	8	7	9	4	2	6	1	3	5
成本比重 (%)	39	28	15	4	3	3	3	2	2	1
累积比重 (%)	67		19			14				
类别	A		B			C				

计算结果列于表 36.1-17 并把它们分成 ABC 三类。从而决定将零件 8 及 10 定为优化对象。

3) 功能成本双重比率分析法 这个方法是综合了成本分析和功能分析, 以功能成本比重的平均值作为标准, 选择功能重要、成本高, 从而平均值高者作为优化对象。

实例分析:

某项目的制造费用主要用于机械、控制、软件和监测四部分, 为了降低成本必须对其进行优化。其具体情况列于表 36.1-18。

表 36.1-18 功能成本双重比率分析法

项目	功能得分	功能比重 (%)	目前费用 / 千元	成本比重 (%)	功能成本平均比 (%)	选择次序
机械	90	28	100	25	26.5	2
控制	80	25	50	12.5	18.8	3
软件	70	22	50	12.5	17.3	4
监测	80	25	200	50	37.5	1

表中:

$$\text{功能比重} = \frac{\text{项目功能得分}}{\text{总功能得分}}$$

$$\text{成本比重} = \frac{\text{项目成本}}{\text{总成本}}$$

$$\text{功能成本平均比重} = \frac{\text{功能比重} + \text{成本比重}}{2}$$

从表 36.1-18 可见, 监测项功能与成本平均比重最高, 反映出该项目的重要性与耗费总体指标最高, 应选为优化对象。

(3) 从市场分析角度选择优化对象

用以下方法从市场分析角度选择优化对象。

1) 产品寿命周期法 任何一种产品都有其寿命周期。根据不同的销售量和盈利状况可划分投入期、成长期、成熟期和退让期。寿命周期法就是通过分析产品所处的寿命期阶段来选择优化对象, 并确定优化目标。它

认为: 当产品在市场上逐步进入退让期时, 企业应及时对产品设计进行优化, 投入新技术与新功能使其成为升级换代的新产品。

2) 用户评估法 在市场上通过向用户发放调查表、征询顾客对产品的意见和建议。根据顾客的意见确定以产品的某项功能或某一方面为优化对象。

如: 20 世纪 80 年代初, 美国某公司剃须刀产品市场, 受到竞争对手生产出一种新型的具有可更换刀片机构的剃须刀的威胁。于是该公司组织了用户调查, 调查发现用户在选择剃须刀时首先考虑的是安全, 而不是易于更换刀片的问题。该公司决定不跟随竞争对手之后开发具有易于更换刀片机构的新产品, 而是重点提高剃须刀使用的安全性。实践证明, 该项决策是正确的。

(4) 功能评价指数法

综合产品的技术、经济等指标, 产生了一些从价值角度评价产品的各个环节, 进而确定优化对象的方法。这种方法首先求出各环节的功能指数和成本指数, 然后求出价值指数, 并且定义价值指数 =  $\frac{\text{功能指数}}{\text{成本指数}}$ 。最后根据价值指数的大小来选择优化对象。其具体方法有以下几种:

1) 强制确定法, 亦称 FD 法 强制确定法 (Forced Decision Method), 是先求出功能指数、成本指数和价值指数, 根据价值指数选择优化对象。然后还可以按功能指数分配预计成本, 求出成本降低指标, 以实现功能价值的最优化。具体做法是:

① 求功能指数 将要分析的方案各个环节列表, 按各自功能的重要程度作一对一比较, 功能重要得 1 分, 功能次要得 0 分, 依次下去, 计算出累计得分, 就可求得某环节的功能指数, 其计算公式为

$$F_i = \frac{f_i}{\sum_{i=1}^n f_i} \quad (i = 1, 2, \dots, n)$$

式中  $F_i$  —— 第  $i$  个环节的功能指数;

$f_i$  —— 第  $i$  个环节的累计得分;

$n$  —— 某方案的环节数。

② 求成本指数 计算公式为

$$C_k = \frac{C_i}{\sum_{i=1}^n C_i} \quad (i = 1, 2, \dots, n)$$

式中  $C_k$  —— 成本指数;

$C_i$  —— 各环节的现实成本。

可通过调查, 得到各环节的现实成本。

③ 求价值指数 价值指数等于功能指数比成本指数。

$$V_i = \frac{F_i}{C_k} \quad (i = 1, 2, \dots, n)$$

式中  $V_i$  价值指数,

① 价值指数的分析 根据  $V_i$  的不同数值, 作如下分析:

$V_i=1$ , 即功能与成本所占比重一致, 属非优化对象;

$V_i=0$ , 说明该环节没有价值, 可以合并或省去;

$V_i>1$ , 表明功能较重要, 或是运用了先进技术使得成本较低; 若  $V_i$  值过大, 则可能是成本太低或投入不足;

$V_i<1$ , 表明功能不重要, 或成本偏高。

根据国外经验,  $V_i>3$  以上或  $V_i<0.5$  以下的, 均可作为重点分析和改进对象, 即可选定为须优化项目。

⑤ 确定目标成本 经过价值分析, 可发现有些环节的功能与成本之间不协调, 需根据计算所得功能指数, 重新分配现实成本。计算公式是

$$C_i = F_i \times \sum_{i=1}^n C_i$$

式中  $C_i$ ——重新分配的现实成本。

为降低成本, 可根据调查和预测, 得到一个总控制成本  $C_p$ , 然后按功能指数  $F_i$  的大小分摊到各环节上去, 作为各环节的控制成本, 即目标成本。公式如下

$$C_i^0 = F_i \times C_p$$

式中  $C_i^0$  各环节的目标成本。

求得  $C_i^0$  后, 就可得到通过优化以后的各环节预期的成本降低幅度, 即  $C_i - C_i^0$ 。

实例分析:

某方案有 8 个环节, 已知各自的现实成本, 现用 FD 法进行分析。设  $C_p = 5000$  元。

解:

1. 用“01”评分法进行一对一比较, 求出各环节的功能指数, 列于表 36.1-19。

表 36.1-19 “01”评分及功能指数

环节	一对一比较结果								得分	功能系数
	A	B	C	D	E	F	G	H		
A	×	1	1	0	1	1	1	1	6	0.214
B	0	×	1	0	1	1	1	1	5	0.179
C	0	0	×	0	1	1	1	0	3	0.107
D	1	1	1	×	1	1	1	1	7	0.250
E	0	0	0	0	×	0	1	0	1	0.036
F	0	0	0	0	1	×	1	0	2	0.071
G	0	0	0	0	0	0	×	0	0	0
H	0	0	1	0	1	1	1	×	4	0.143
$\Sigma$									28	1

2. 已知的现实成本  $C_i$ , 利用公式分别求得成本指数  $C_i^0$  和价值指数  $V_i$ , 列于表 36.1-20。

3. 求重新分配的现实成本和改进后的目标成本, 最终看出经过优化后成本的降低幅度, 列于表 36.1-21。

表 36.1-20 现实成本、成本指数、价值指数表

环节	$F_i$	$C_i$ (元)	$C_i^0$	$V_i$
A	0.214	1818	0.2796	0.77
B	0.179	2800	0.3537	0.51
C	0.107	280	0.043	2.19
D	0.250	278	0.043	5.81
E	0.036	620	0.0953	0.38
F	0.071	480	0.0738	0.96
G	0	75	0.0115	0
H	0.143	651	0.1001	1.43
$\Sigma$	1	6502	1	1

表 36.1 21 分摊后现实成本、目标成本及成本降低幅度

环节	$C_i$ (元)	$C_i^0$ (元)	$C_i - C_i^0$ (降低成本幅度)
A	1391	1070	748
B	1164	895	1405
C	696	535	-255
D	1626	1250	-972
E	234	180	440
F	462	355	125
G	0	0	75
H	929	715	-64
$\Sigma$	6502	5000	1502

2) 理想区域法 这个方法是由日本田中教授提出的。思路是: 价值指数相同的对象, 由于各自的成本指数及功能指数的绝对值不同, 因而可能对方案的实际影响有很大差异。所以在选择优化对象时, 不应把价值指数相同的对象同等看待。而应选择对设计方案实际影响大的环节。

① 基本原理 已知价值指数  $V_i = \frac{F_i}{C_i^0}$ , 如果用  $y_i$  表示  $F_i$ ,  $x_i$  表示  $C_i^0$ , 则以  $Y$  为纵轴,  $X$  为横轴画坐标图, 如图 36.1-23 所示。通过计算来确定理想区域, 在区域内的点不作为重点分析对象, 在区域外的点所代表的环节作为重点分析对象。已知  $V_i$  的比值以稍大于 1 比较理想, 因此, 理想区域应是围绕  $V_i=1$  的某个区域。

② 理想区域的确定  $V_i=1$  在图 36.1-23 上是一条 45° 线。构成理想区域的点  $P_i$  应离散地分布于 45° 线两侧。当  $P_i$  点距离原点  $O$  远时, 说明  $y_i$  和  $x_i$  的绝对值大, 对整体的实际影响也大, 故允许偏离 45° 线的值  $r$  要大些; 反之,  $P_i$  点距原点  $O$  近, 则影响小, 允许的  $r$  值可大些,  $O$  到  $P_i$  点距离用矢量  $OP$  表示,  $O$  点到  $Q$  点距离用矢量  $i$  表示。且  $r \perp i$ , 如图 36.1-24 所示。理

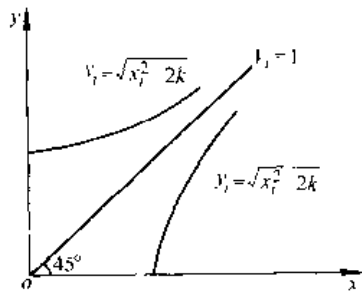


图 36.1-23 理想区域

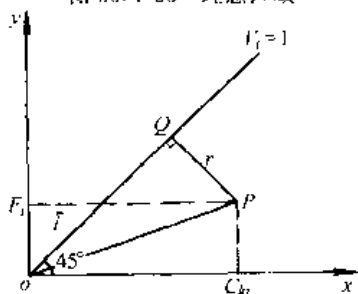


图 36.1-24 理想区域的确定

理想区域的界限按  $r$  与  $i$  的乘积为常数，即  $ri=k$  确定，理想区域上、下两边缘曲线方程为（证明略）

$$\begin{cases} y_1 = \sqrt{x_1^2 - 2k} \\ y_2 = \sqrt{x_1^2 + 2k} \end{cases}$$

常数  $k$  可通过调查或试验得到。

实例分析：

已求得方案各环节的价值指数，功能指数和成本指数见表 36.1-19。试用理想区域法确定优化对象。

解 通过试验，取  $k=0.02$ ，在价值指数坐标图上做出： $y = \sqrt{x^2 \pm 0.04}$  的两条曲线，则得到理想区域， $y_1$  为功能指数  $F_i$ ， $x_1$  为成本指数  $C_k$ 。这样，可一一标出各环节在理想区域中的位置，如图 36.1-25 所示。从中可见  $B$  环节和  $D$  环节都在理想区域外，故应作为重点优化对象。

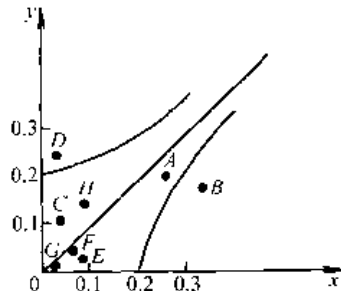


图 36.1-25 用理想区域法确定优化对象

### 5.3 总体设计的评价

(1) 基本术语

1) 评价目标 评价目标见表 36.1-22。

表 36.1-22 评价目标

技术评价目标	性能指标、工艺性、可靠性……
经济评价目标	成本、利润、投资回收期……
社会评价目标	对社会的影响：节能、环境污染……

评价项目一般为 6~10 项。

2) 加权系数 定量评价时，根据目标的相对重要性设置加权系数。一般取各评价目标的加权系数均小于 1，而全部加权系数之和等于 1。

确定加权系数的判别表法：两两比较评价目标的重要程度，同等重要为 2:2；不等则为 3:1 或 4:0。列表计算出各加权系数  $g_i$ 。

$$g_i = k_i / 2 (n^2 - n)$$

式中  $k_i$  - 各评价目标的总分；

$n$  - 评价目标数。

3) 评价目标树 总目标分解为不同一级子目标及各自的加权系数。一级子目标分解为二级子目标及各自的加权系数……，最后形成评价目标树。

(2) 评价方法

评价方法见表 36.1-23。

表 36.1-23 评价方法

简单评价法	点评价法	各方案按评价目标逐项作初步评价，用可行 (+)、不行 (-)、信息不足 (?) 表示，最后总评
	名次计分法	$m$ 个专家评价 $n$ 个方案，每人对每一方案排出名次，最高 $n$ 分，最低 1 分，最后把每个方案得分相加得每个方案总分 $x_i$ 。专家们意见的一致性用一致性系数 $c$ 表示： $c = \frac{12 \left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]}{m^2 (n^3 - n)}$
评分法	直接计值法	根据 $n$ 个评价目标的允许分值、要求分值和理想分值分别定为 0、8、10 (10 分制) 或 0、4、5 (5 分制)；再根据 $m$ 个方案各评价目标的具体参数值线性插入，求得相对应的评分值 $p_i$ 。 $m$ 个方案的总分值： 相加总分值 $Q_1 = \sum_{i=1}^m p_i$ (计算简单) 连乘总分值 $Q_2 = \prod_{i=1}^m p_i$ (各方案总分值差异大，便于比较) 平均总分值 $Q_3 = \frac{1}{n} \sum_{i=1}^m p_i$ (计算简单直观) 相对总分值 $Q_4 = \frac{\sum_{i=1}^m p_i}{nQ_0} \leq 1$ ( $Q_0$ 为理想方案总分值，能看出与理想方案的差距)

(续)

<p>评分法</p>	<p>有效值法</p>	<p>确定 <math>n</math> 个评价目标的加权系数矩阵 <math>G</math></p> $G = (g_1, g_2, \dots, g_n) \quad g_i < 1, \sum g_i = 1, i = 1 \sim n$ <p><math>m</math> 个方案对 <math>n</math> 个评价目标的评分值矩阵 <math>P</math></p> $P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_i \\ \vdots \\ P_m \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \dots & p_{in} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mn} \end{bmatrix}$ <p><math>m</math> 个方案的有效值矩阵 <math>N</math></p> $N = GP^T = [N_1, N_2, \dots, N_i, \dots, N_m]$ <p>第 <math>j</math> 个方案的有效值 <math>N_j</math></p> $N_j = GP_j^T = g_1 p_{j1} + g_2 p_{j2} + \dots + g_n p_{jn}$ <p>比较各方案的有效值, 最大者为最佳方案</p>
<p>技术经济评价法</p>	<p>相对价法</p>	<p>某方案的技术价 <math>W_t</math>:</p> $W_t = \frac{\sum_{i=1}^n p_i g_i}{p_{max}} \leq 1 \quad (p_{max} \text{ 为最高分值})$ <p>一般 <math>W_t &lt; 0.6</math> 为不合格</p> <p>某方案的经济价 <math>W_w</math>, 由实际生产成本 <math>H</math>、理想生产成本 <math>H_I</math>、允许生产成本 <math>H_p</math> 决定</p> $H_I \approx 0.7 H_p$ $W_w = \frac{H_I}{H} \leq 1$ <p>一般 <math>W_w &lt; 0.7</math> 为不合格</p> <p>某方案相对价 <math>W</math></p> $W = \sqrt{W_t W_w}$ <p>一般 <math>W &lt; 0.65</math> 为不合格</p>
<p>优度图法</p>		<p>以技术价 <math>W_t</math> 与经济价 <math>W_w</math> 构成平面坐标系, <math>W_t \geq 0.6, W_w \geq 0.7</math> 为许用区。可看出各方案的技术—经济综合性能, 便于提出改进方向</p>
<p>单评价目标</p>		<p>用集合与模糊数学, 将模糊信息数值化, 进行定量评价</p> <p>统计法求隶属度: 例如一群顾客中, 对某产品性能的评价 (优、良、中、差) 的百分数分别为 <math>x_1\%</math>、<math>x_2\%</math>、<math>x_3\%</math>、<math>x_4\%</math>, 则隶属度组成的模糊评价集 <math>x</math> 为:</p> $x = \{x_1, x_2, x_3, x_4\} = \{\text{优、良、中、差}\}$ <p>由隶属函数求隶属度: 选择模糊数学中合适的典型隶属函数, 求得特定条件下的隶属度</p>
<p>模糊评价法</p>	<p>多评价目标</p>	<p><math>n</math> 个评价目标, <math>m</math> 个模糊评价, <math>n</math> 个加权系数</p> <p>评价目标集 <math>Y = \{y_1, y_2, \dots, y_n\}</math></p> <p>评价集 <math>X = \{x_1, x_2, \dots, x_m\}</math></p> <p>加权系数集 <math>G = \{g_1, g_2, \dots, g_n\} \quad (\sum_{i=1}^n g_i = 1)</math></p> <p><math>z</math> 个方案中第 <math>k</math> 个方案对 <math>n</math> 个评价目标的模糊评价矩阵 <math>R_k</math></p> $R_k = \begin{bmatrix} R_{k1} \\ R_{k2} \\ \vdots \\ R_{ki} \\ \vdots \\ R_{kn} \end{bmatrix} = \begin{bmatrix} r_{k1} & r_{k2} & \dots & r_{k1n} \\ r_{k2} & r_{k2} & \dots & r_{k2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{ki} & r_{i2} & \dots & r_{in} \\ \vdots & \vdots & \ddots & \vdots \\ r_{kn} & r_{n2} & \dots & r_{nm} \end{bmatrix} \quad (k=1 \sim z)$ <p>各方案加权的综合模糊评价</p> $B_k = GR_k = [b_1, b_2, \dots, b_j, \dots, b_m] \quad (k=1 \sim z)$ <p>求 <math>b_j</math> 的方法有:</p> <p>1. <math>M(\wedge, \vee)</math>: 按取小 (<math>\wedge</math>) 取大 (<math>\vee</math>) 运算合成矩阵</p> $g \wedge b = \min(g, b)$ $g \vee b = \max(g, b)$ $b_j = \bigvee_{i=1}^n (g_i \wedge r_{ij})$ <p>2. <math>M(\cdot, +)</math>: 按乘 (<math>\cdot</math>) 加 (<math>+</math>) 运算进行矩阵合成</p> $b_j = \sum_{i=1}^n g_i r_{ij}, \quad j = 1, 2, \dots, m$ <p>对各 <math>B_k</math> 值进行归一化处理, 折算为按百分比表示的隶属度, 按最大隶属度原则判断方案优劣顺序</p>

(3) 评价方法应用实例

1) 评分法应用实例 某厂开发一电池产品, 技术指标 2 项, 即: 额定电压, 有效寿命; 经济指标一项, 即电池生产成本。表 36.1-24 列出这 3 项指标的理想值、优等值和及格值, 并分别给出 3 项目标的加权系数。通过对 3 种方案样品的实际考核, 得出表中的实际值。相应的评价目标树如图 36.1-26 所示。

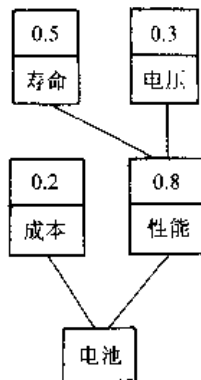


图 36.1-26 电池的评价目标树

由 3 项目标的理想值、优等值和及格值可以作出各自的评分线图, 如图 36.1-27a、b 及 c 所示。这里约定及格分为 1, 中等为 2, 良好为 3, 优等分为 4, 理想分为 5。从图中可以查得各实际值的得分, 列于表 36.1-24。

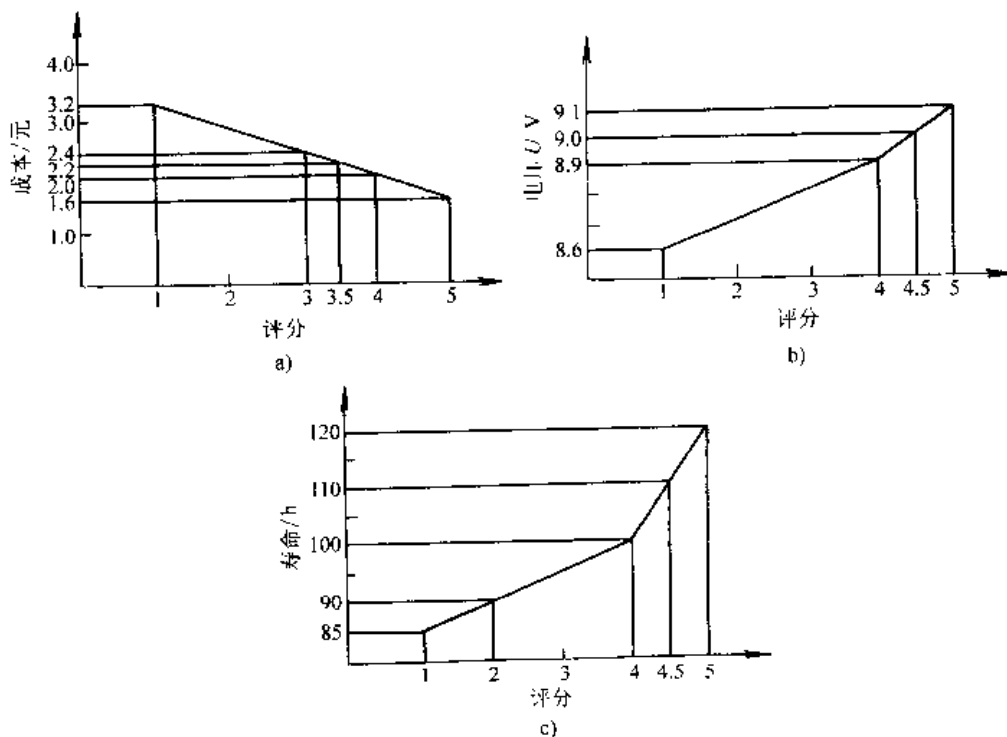


图 36.1-27 电池的评分线图

a) 成本评分线图 b) 电压评分线图 c) 寿命评分线图

表 36.1-24 电池的评价目标值

评价目标	成本/元	电压/V	寿命/h	
理想值	1.6	9.1	120	
优等值	2	8.9	100	
及格值	3.2	8.6	85	
加权系数	0.2	0.3	0.5	
实际值	A	2.4	9	100
	B	2.2	8.6	110
	C	1.6	8.9	90

在 5 种总分计算法中, 分值相加法计算最简单; 分值连乘法可使各方案总分差值加大, 便于比较; 平均总分值比较直观, 相对总分值为平均总分与理想分之比, 能看出与理想方案的差距, 加权分值相加法则考虑了权重。

表 36.1-25 列出 5 种评价参考总分, 显然, 方案 A 为最佳方案, 若能改进方案 B 的电压指标, 方案 C 的寿命指标, 则可进行新一轮评选。

2) 技术经济评价法应用实例 一高速专用磨床的主轴前支承被列为优化对象, 共提出三种方案, 即: 采用 3 排深沟球轴承, 3 排液体动压轴承, 或 3 排球面液体静压轴承。列出 3 项技术评价目标: 性能、制造和使用, 在性能之下开列了刚度和抗振性 2 项子目标, 在制

表 36.1-25 电池评分表

评价目标	成本	电压	寿命	总分					
				分值相加	分值连乘	平均分	相对总分	加权分值相加	
加权系数	0.2	0.3	0.5						
方案评分	A	3	4.5	4	11.5	54	3.83	0.77	3.95
	B	3.5	1	4.5	9	15.75	3.00	0.60	3.25
	C	5	4	2	11	40	3.67	0.74	3.2

造之下并列了加工与装配 2 项子目标, 其技术评价目标树如图 36.1-28 所示。图中同时给出了加权系数值。

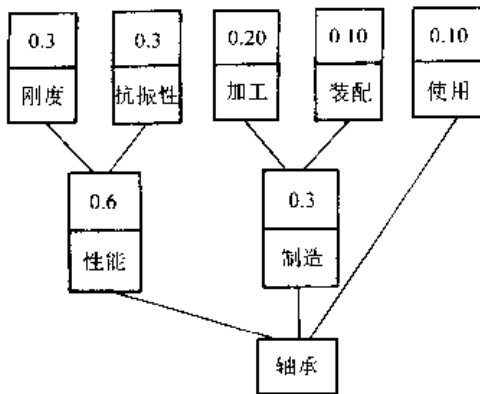


图 36.1-28 轴承方案的技术评价目标树

表 36.1-26 列出 3 种轴承的技术评价表。请专家分别对 3 种方案按照 5 个具体目标打分, 并规定最高分为 5 分。由此得每一方案考虑权重后的得分  $\Sigma b_j q_j$

$$\sum_{j=1}^5 b_{ij} q_j (i=1, 2, 3)$$

各方案的技术价

$$W_i = \frac{\sum_{j=1}^5 b_{ij} q_j}{b_{max}}$$

式中  $b_{max}$  一约定最高打分, 这里  $b_{max}=5$ 。

3 种方案的技术价均列于表 36.1-26 中。

表 36.1-26 轴承方案的技术评价

评价目标	方案 加权系数 $q_j$	滚动		动压		静压	
		$b_{1j}$	$b_{1j} q_j$	$b_{2j}$	$b_{2j} q_j$	$b_{3j}$	$b_{3j} q_j$
1. 刚度	0.3	4	1.2	3.5	1.05	3	0.9
2. 抗振性	0.3	2	0.6	3	0.9	4	1.2
3. 加工	0.2	4	0.8	3	0.6	2.5	0.5
4. 装配	0.1	4	0.4	3	0.3	2.5	0.25
5. 使用	0.1	2	0.2	3.5	0.35	4.5	0.45
$b_{max}=5$		$\Sigma b_{1j} q_j=3.2$		$\Sigma b_{2j} q_j=3.2$		$\Sigma b_{3j} q_j=3.3$	
$W_T = \Sigma b_{ij} q_j / b_{max}$		$W_{T1}=0.64$		$W_{T2}=0.64$		$W_{T3}=0.66$	

表 36.1-27 列出了各方案的生产成本  $H_i$ , 以及设计者所希望达到的理想成本  $H_1$ , 由此得出各方案的经

表 36.1-27 轴承方案的经济评价

方案	滚动	动压	静压
评价目标			
生产成本 $H_i$ (%)	140	160	180
理想成本 $H_1$ (%)	100		
经济价 $W_w = H_i / H_1$	$W_{w1}=0.714$	$W_{w2}=0.625$	$W_{w3}=0.555$

济价  $W_w$

$$W_{wi} = \frac{H_i}{H_1} (i=1, 2, 3)$$

根据各方案的技术、经济价, 可得其相对价

$$W_1 = \sqrt{W_{T1} W_w} = \sqrt{0.64 \times 0.714} = 0.676$$

$$W_2 = \sqrt{W_{T2} W_w} = \sqrt{0.64 \times 0.625} = 0.632$$

$$W_3 = \sqrt{W_{T3} W_w} = \sqrt{0.66 \times 0.555} = 0.605$$

一般  $W_T < 0.6$  为不合格,  $W_w < 0.7$  为不合格,  $W < 0.65$  为不合格。从优度图 36.1-29 可见, 只有方案 1 落在许用的区域内 (阴影区)。

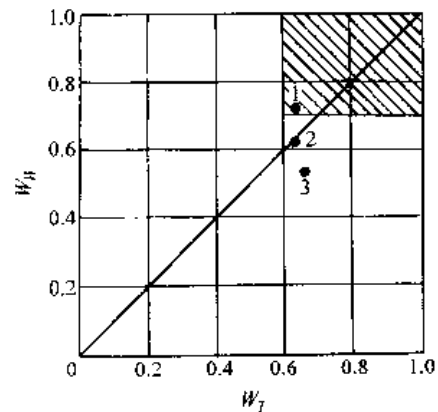


图 36.1-29 优度图

3) 模糊评价法 今确定一计算机管理系统为优化对象, 提出了购买专用软件, 购买通用软件或自行开发 3 种方案, 投资必须控制在 30 万元以内, 15 万元为中, 低于 5 万元为优, 同时建立性能和周期等评价目标, 以及相应的加权系数, 评价目标树如图 36.1-30 所示, 专家按优、中、差三级打分的结果列于表 36.1-28。

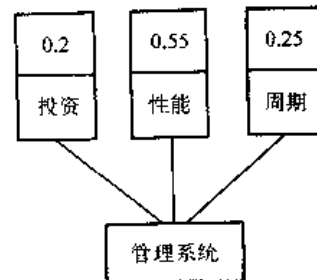


图 36.1-30 管理系统的评价目标树

按照模糊评价的法则, 依次进行如下工作:

- ① 建立评价目标集

表 36.1-28 计算机管理系统的目标情况

评价目标	投资 (1)	性能 (2)	周期 (3)
I 购买专用软件	28	优	短
II 购买通用软件	12	中	中
III 自行开发	6	差	长

$$B = \{b_1, b_2, \dots, b_n\} \quad (n = \text{评价目标数})$$

本例中为  $B = \{\text{投资, 性能, 周期}\}$

### ② 建立加权系数集

$$Q = \{q_1, q_2, \dots, q_n\} \quad (n = \text{评价目标数})$$

本例中为  $Q = \{0.2, 0.55, 0.25\}$

### ③ 建立模糊评价集

$$P = \{p_1, p_2, \dots, p_m\} \quad (m = \text{评价数})$$

本例中为  $P = \{\text{优, 中, 差}\}$

### ④ 进行模糊评判 请一组专家, 对评价目标集

中每一目标  $b_i (i = 1, 2, \dots, n)$  按照评价集中的评价打勾 (“√”)。这项工作针对每一项方案进行, 对于某一评价目标  $b_i$ , 有  $z$  个评判结果

$$R_{1i} = \{Y_{11}, Y_{21}, \dots, Y_{m1}\}$$

$$R_{2i} = \{Y_{12}, Y_{22}, \dots, Y_{m2}\}$$

$$R_{zi} = \{Y_{1z}, Y_{2z}, \dots, Y_{mz}\} \quad (z = \text{方案数})$$

...

$$R_{ni} = \{Y_{1n}, Y_{2n}, \dots, Y_{mn}\}$$

其中某一元素  $Y_{mz}$  表示在第  $z$  方案中, 对评价目标  $b_i$ , 持  $p_m$  评价的专家占总数的百分比。

在本例中, 评价目标数  $n = 3$ , 评价数  $m = 3$ , 方案数  $z = 3$ 。

对于目标  $b_1$  (投资), 得模糊评判结果

$$R_{11} = \{0, 0.13, 0.87\} \quad \text{方案 I}$$

$$R_{11} = \{0.3, 0.7, 0\} \quad \text{方案 II}$$

$$R_{11} = \{0.9, 0.1, 0\} \quad \text{方案 III}$$

它表明, 对于投资这一评价目标, 在方案 I 中, 持评价为“优”的专家为零, 持意见“中”的专家占总数 13%, 持意见“差”的专家占总数 87%。

对于评价目标  $b_2$  (性能), 得模糊评判结果

$$R_{12} = \{0.95, 0.05, 0\} \quad \text{方案 I}$$

$$R_{12} = \{0.05, 0.80, 0.15\} \quad \text{方案 II}$$

$$R_{12} = \{0.10, 0.30, 0.60\} \quad \text{方案 III}$$

对于评价目标  $b_3$  (周期), 得模糊评判结果:

$$R_{13} = \{0.85, 0.10, 0.05\} \quad \text{方案 I}$$

$$R_{13} = \{0.20, 0.60, 0.20\} \quad \text{方案 II}$$

$$R_{13} = \{0.10, 0.20, 0.70\} \quad \text{方案 III}$$

这里, 模糊评判结果又称为隶属度。

### ⑤ 建立模糊评价矩阵

$$R_i = \{R_{i1}, R_{i2}, \dots, R_{in}\}^T$$

$$R_1 = \{R_{11}, R_{12}, \dots, R_{1n}\}^T$$

$$R_2 = \{R_{21}, R_{22}, \dots, R_{2n}\}^T$$

...

$$R_n = \{R_{n1}, R_{n2}, \dots, R_{nn}\}^T$$

在本例中

$$R_1 = \begin{bmatrix} 0 & 0.13 & 0.87 \\ 0.95 & 0.05 & 0 \\ 0.85 & 0.10 & 0.05 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.05 & 0.80 & 0.15 \\ 0.20 & 0.60 & 0.20 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$$

矩阵中的每一行对应着一个评价指标的隶属度。

⑥ 作出模糊评价 模糊评价矩阵虽反映了评价的全貌, 但没有反映加权系数, 在模糊数学中有两种将模糊评价矩阵与加权系数集合成, 建立综合模糊评价的方法。

第一种: 作模糊矩阵乘法, 这里

$$K = Q \cdot R$$

符号“·”是模糊矩阵乘法运算符号, 其算法与普通矩阵乘法运算相似, 只不过  $Q$  中每一元素不是和  $R$  中某列元素一一相乘然后相加, 而是一一相比取其小值, 然后从获得的一系列小值中取一最大值。在本例中

$$Z_1 = \{0.2, 0.55, 0.25\} \cdot \begin{bmatrix} 0 & 0.13 & 0.87 \\ 0.95 & 0.05 & 0 \\ 0.85 & 0.10 & 0.05 \end{bmatrix} \\ = \{0.55, 0.13, 0.2\}$$

$$Z_2 = \{0.2, 0.55, 0.25\} \cdot \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.05 & 0.8 & 0.15 \\ 0.20 & 0.60 & 0.20 \end{bmatrix} \\ = \{0.20, 0.55, 0.20\}$$

$$Z_3 = \{0.2, 0.55, 0.25\} \cdot \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.2 & 0.7 \end{bmatrix} \\ = \{0.20, 0.3, 0.55\}$$

如  $K_1$  中,  $Q$  与  $R$  中第一列作“·”运算, 0.2 比 0 取 0, 0.55 比 0.95 取 0.55, 0.25 比 0.85 取 0.25, 然后取 0、0.55、0.25 中的大者 0.55 为  $K_1$  中第 1 元素, 以此类推。

为更清楚地进行比较, 将  $K$  作归一处理, 在本例中

$$Z_1 = \frac{\{0.55, 0.13, 0.20\}}{0.55 + 0.13 + 0.20}$$

$$\begin{aligned}
 &= \{0.625, 0.148, 0.227\} \\
 Z_1 &= \frac{\{0.20, 0.55, 0.20\}}{0.20+0.55+0.20} \\
 &= \{0.21, 0.58, 0.21\} \\
 Z_2 &= \frac{\{0.20, 0.30, 0.55\}}{0.20+0.30+0.55} \\
 &= \{0.19, 0.29, 0.52\}
 \end{aligned}$$

它表明,对方案 I,评价为“优”的隶属度为 62.5%，“优”与“中”之和为 77.3%，方案 II 中“优”与“中”之和为 79%，方案 I 与 II “优”“中”之和相近，前者“优”明显领先，故决策 I 为最佳方案。

第二种：将 Q 与 R 按一般矩阵相乘，然后作归一处理，在本例中

$$\begin{aligned}
 &\left. \begin{aligned} 0.2 \times 0 + 0.55 \times 0.95 + 0.25 \times 0.85 &= 0.935 \\ 0.2 \times 0.13 + 0.55 \times 0.05 + 0.25 \times 0.1 &= 0.0785 \\ 0.2 \times 0.8 + 0.55 \times 0 + 0.25 \times 0.05 &= 0.1725 \end{aligned} \right\} \\
 &= \frac{[0.935, 0.0785, 0.1725]}{0.935 + 0.0785 + 0.1725} \\
 &= [0.788, 0.066, 0.146] \\
 &\left. \begin{aligned} 0.2 \times 0.3 + 0.55 \times 0.05 + 0.25 \times 0.2 &= 0.1375 \\ 0.2 \times 0.7 + 0.55 \times 0.8 + 0.25 \times 0.6 &= 0.73 \\ 0.2 \times 0 + 0.55 \times 0.15 + 0.25 \times 0.2 &= 0.1325 \end{aligned} \right\} \\
 &= [0.1375, 0.73, 0.1325] \\
 &\left. \begin{aligned} 0.2 \times 0.9 + 0.55 \times 0.1 + 0.25 \times 0.1 &= 0.26 \\ 0.2 \times 0.1 + 0.55 \times 0.3 + 0.25 \times 0.2 &= 0.235 \\ 0.2 \times 0 + 0.55 \times 0.6 + 0.25 \times 0.7 &= 0.505 \end{aligned} \right\} \\
 &= [0.26, 0.235, 0.505]
 \end{aligned}$$

得  $Z_1 = [0.788, 0.066, 0.146]$   $Z_2 = [0.1375, 0.73, 0.1325]$   $Z_3 = [0.26, 0.235, 0.505]$

显然仍是方案 I 为最佳。

### 5.4 总体设计的优化

根据对总体设计中关键部位评价的结果，找出问题之所在，然后运用创新思维原理与方法，寻求该部位的优化解，达到总体设计优化的目的，也可以在设计之初就建立几种方案，然后通过评价，选出最佳方案。

#### 总体方案优化实例

这里将燕山大学所作碱性电池自动涂胶机系统最佳方案设计作为应用实例予以介绍。

碱性电池的涂封口胶工艺，是生产碱性电池的重要环节，目前大多数为手工操作，为保证质量，提高生产率，须要开发一台自动涂胶机。

1) 总体方案设计 表 36.1-29 列出了自动涂胶机的设计要求，其总功能是将封口胶自动准确地涂敷在壳口的内圆周表面上，用黑箱理论描述如图 36.1-31 所示。

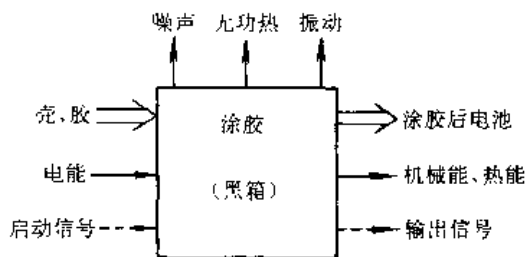


图 36.1-31 涂胶机的技术系统与环境的黑箱模型

表 36.1-29 自动涂胶机设计要求

要求	数值	重要性
必达条件		
涂胶均匀准确		I
钢壳不得划伤损坏		
机器工作可靠、运转平稳		
机器生产率	40.60 只/min	
允许制造费用	<2 万元	
最低要求		
操作高度	850~900mm	I
3m 距离内噪声强度	≤70dB	II
大修周期	10 000 工作小时	IV
自动控制胶体温度		IV
希望		
占地面积小		
安装、操作、维修简单		
运转费用低		
设备造价低		
采用机械传动与控制		

I ~ IV -- 重要程度 IV —— 最重要

将总功能分解，在“发散”求解的基础上，形成涂胶机解的形态学矩阵，列于表 36.1-30。

表 36.1-30 涂胶机解的形态学矩阵

功能元	功能元解		
	1	2	3
A 钢壳贮存供送	料斗拨轮	料斗传送带	人工摆放
B 传送钢壳	双曲柄平动	间歇转盘	顶杆
C 涂胶方式	涂胶轮	涂胶嘴	毛刷
D 取胶方式	柱塞泵	胶轮	
E 胶体加热	电加热	蒸汽热	
F 传动方式	机械	液压	气动
G 动力源	电动机	液动机	伺服电动机
H 控制方式	计算机	单片机	运动分配轴
I 钢壳旋转	带摩擦	橡胶轮	绳摩擦
J 钢壳输出	拨爪	顶杆	轨道
	3	2	1

总功能分解为 10 个功能元，为了得到较多的设计方案，应先“发散”，寻求实现功能元的各种可能解。利用形态学矩阵进行方案组合，可形成 N 个组合方案



$$N=3 \times 3 \times 3 \times 2 \times 2 \times 3 \times 3 \times 3 \times 3=26244$$

根据设计要求 (表 36.1-28), 对众多方案进行“收敛”优选, 把矩阵表中与设计不相容的方案剔除, 并根据实际经验, 选择切实可行的解, 组成有价值的 3 个组合方案。

原理方案 I

$$A_1-B_2-C_2-D_1-E_1-F_1-G-H_2-I_2-J_3$$

原理方案 II

$$A_2-B_1-C_1-D_2-E_2-F_1-G_1-H_3-I_1-J_3$$

原理方案 III

$$A_3-B_3-C_1-D_2-E_1-F_1-G_1-H_3-I_1-J_1$$

2) 最优方案选择 对涂胶机系统评价的依据是评价目标。根据设计要求确定 10 项评价目标, 利用互比判别法, 按 4 级重要程度, 通过各项目间一一对比打分, 确定相应的加权系数  $q_i$ , 列于表 36.1-31, 表中  $k_i$  为各目标的得分之和。

方案评分采用 10 分制, 将 4 人评审小组打出的平均分作为评分值, 列于表 36.1-32。

表 36.1-31 加权系数判别计算表

评价目标	比较目标											$k_i = \sum_{j=1}^n k_{ij}$
	涂胶质量	供料通畅性	结构简单可靠	加工维修	生产率	价格	寿命	占地面积	自动化	噪声	$k_i$	
涂胶质量	—	3	3	4	3	3	3	4	3	4	30	0.167
供料通畅性	1	—	3	3	3	3	3	4	3	4	27	0.15
结构简单可靠	1	1	—	2	1	2	2	2	2	3	16	0.089
加工维修	0	1	2	—	1	2	2	2	2	3	15	0.083
生产率	1	1	3	3	—	3	2	3	2	3	21	0.117
价格	1	1	2	2	1	—	2	2	2	3	16	0.089
寿命	1	1	2	2	2	—	—	3	2	3	18	0.10
占地面积	0	0	2	2	1	2	1	—	1	3	12	0.067
自动化	1	1	2	2	2	2	2	3	—	3	18	0.10
噪声	0	0	1	1	1	1	1	1	1	—	7	0.039
$\Sigma$											180	1

表 36.1-32 涂胶机的技术评价

评价因素	权重	性能特征	计量单位	方案 I			方案 II			方案 III		
				性能	评分 $p_{1i}$	$g_i p_{1i}$	性能	评分 $p_{2i}$	$g_i p_{2i}$	性能	评分 $p_{3i}$	$g_i p_{3i}$
1 涂胶质量	0.167	涂层均匀一致	—	超目标	9	1.503	满意	5	0.845	满意	5	0.845
2 供料通畅	0.15	无下壳损伤	—	超目标	9	1.35	很好	8	1.20	很好	8	1.2
3 结构简单可靠	0.089	零件数量少	个	少	7	0.623	少	6	0.534	较少	8	0.712
4 加工维修	0.083	容易加工维修	—	良	7	0.581	满意	6	0.498	良	7	0.581
5 生产率	0.117	涂胶数量	只/h	多	8	0.936	满意	5	0.585	可用	4	0.468
6 价格	0.089	制造成本	万元	低	8	0.712	低	8	0.712	超目标	9	0.747
7 寿命	0.10	工作时间	h	长	8	0.80	短	5	0.50	短	5	0.50
8 占地面积	0.067	占地面积	m <sup>2</sup>	少	8	0.536	少	8	0.536	少	8	0.536
9 自动化	0.10	自动化程度	—	高	8	0.80	高	8	0.80	低	4	0.40
10 噪声	0.039	3m 内噪声	dB	低	8	0.312	低	8	0.312	低	8	0.312
$\Sigma g_i = 1 \quad p_{max} = 10$				$\Sigma g_i p_{1i} = 8.153$	$\Sigma g_i p_{2i} = 6.522$			$\Sigma g_i p_{3i} = 6.301$				
$W_{T1} = \Sigma g_i p_{1i} / (p_{max} \Sigma g_i)$				$W_{T1} = 0.8153$	$W_{T2} = 0.6522$			$W_{T3} = 0.6301$				

自动涂胶机的技术评价通过求方案的技术价  $W_T$  进行。技术价

$$W_T = \frac{\sum_{i=1}^n p_i g_i}{p_{max} \sum_{i=1}^n g_i} = \frac{\sum_{i=1}^n p_i g_i}{p_{max}}$$

式中  $p_i$  ——各技术评价指标的评分值;

$g_i$  ——各技术评价指标的加权系数,  $\sum_{i=1}^n g_i = 1$ ;

$p_{max}$  为最高分值 (取 10 分)。

每个方案的技术价见表 36.1-32。

经济评价通过求方案的经济价  $W_w$  进行, 列于表

36.1-33。用相对成本值代入计算, 求得 3 个方案的经济价。

3 个方案的技术-经济总价值为

表 36.1-33 自动涂胶机设计方案经济评价

经济指标	方案		
	I	II	III
生产成本 $H$ (%)	110	115	100
理想成本 $H_I$ (%)	80	80	80
经济价 $W_w = \frac{H_I}{H}$	0.727	0.696	0.8

$$W_1 = \sqrt{W_{T1} W_{w1}} = \sqrt{0.8153 \times 0.727} = 0.77$$

$$W_2 = \sqrt{W_{T2} W_{w2}} = \sqrt{0.6522 \times 0.696} = 0.674$$

$$W_3 = \sqrt{W_{T3} W_{w3}} = \sqrt{0.6301 \times 0.8} = 0.71$$

3 个方案的优度图如图 36.1-32 所示。由图可知，方案 I ~ III 总价均大于 0.65，可以满足要求，图中阴影区为许用区，方案 II 在区外，方案 I 离开发线最近，定为最优方案。

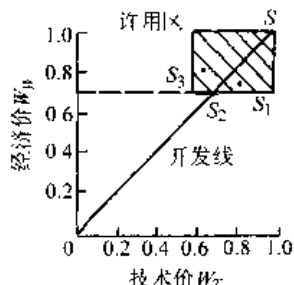


图 36.1-32 3 个方案的优度图

## 6 方案设计和技术设计案例分析

本节以一个家用单手把混水器(冷热水龙头)为例子,从中说明方案设计的步骤和实际应用。还有一个冲击试验台的完整例子,说明了方案设计阶段和技术设计阶段的进程。

### 6.1 家用单手把混水器方案设计案例分析

单手把混水器只需用一个手把就能既调节水温又调节水量。这个任务一般是由产品计划部门向设计部门提出的。

需求说明:需开发一种家用单手把混水器,其数据如下:

- 流通量 10 l/min
- 最大压力 0.6MPa
- 正常压力 0.2MPa
- 热水温度 60°C
- 接头尺寸 1/2in

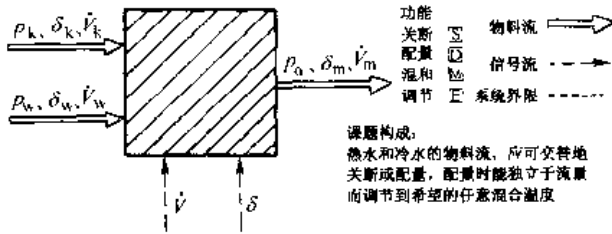
开发的产品应在两年内进入市场,每月约 3000 件时,制造成本不得超过 30 马克。

工作步骤 1:阐明任务书,制订要求表 根据管道连接关系、适用的标准和规定,以及人机工程方面的信息,在修改一张初步的要求表后,得图 36.1-33 所示的第 2 版本。

		用 于 单 手 把 混 水 器		张: 1 页: 1
改动	必/愿	要 求		负责者
	必	1	流通量(混合流):当附件间的压力为 0.2MPa 时,最大流通量为 10 l/min	
	必	2	压力:最大为 1MPa(试验压力 1.5MPa),按 DIN2401)	
	必	3	水流:正常为 60°C,最大为 100°C(短时间的)	
	必	4	温度调节独立于流通量和压力	
	必	5	热水和冷水管压力热±0.05MPa 时的容许温度偏差:±5°C	
	必	6	接头:2×Cu-管 10×1mm, L=400mm	
	必	7	单孔固装 φ35±0.1mm,水盆穿透高度 0 到 18mm.注意水盆尺寸(DIN EN31, DIN EN32, DIN1358)	
	必	8	流出高度,高于水盆上缘 50mm	
	必	9	作为水盆附件来解	
	必	10	可改装为安装附件	
	必	11	操作力小(小孩)(Rohrmeft, W., Hettinger, Th., Kötperkräfte im Rewegungsraum. Berlin 1963)	
	必	12	无外来能量	
	必	13	水的状态是含钙的(注意饮用水质量条件)	
	必	14	温度调节是明确可以知道	
	必	15	厂牌装设得醒目	
	必	16	调停时两个水支路不发生短路	
	必	17	放水时不发生短路	
	必	18	手把只准热到-35°C	
	必	19	碰到附件时不会灼伤	
	必	20	如增费不多,则装设防烫装置	
	必	21	意义明确的操作,简单而舒适的使用(Roomert, W., Arbeitswiss. Prüfliste Zur Arbeitsgestaltung, Berlin 1966)	
	必	22	光滑,易清洁的外廓,无尖锐棱角	
	必	23	运用时噪声低(噪声水平 $I_{Ae} \leq 20dB(A)$ ,测量按 DIN52218)	
	必	24	寿命:10 年,约使用 300000 次	
	必	25	混水器维护容易,维修简单,采用外购配件	
	必	26	最高制造成本:30 马克(3000 件/月)	
	必	27	从开始开发算起的期限:	
			方案设计/详细设计/施工设计/样品 2      4      6      9      个月后	
			代替 5.12 第 1 版	

图 36.1-33 单手把混水器的要求表

工作步骤2: 抽象化, 识别实质性问题 抽象化的基础是要求表, 其结果是图 36.1-34。由于家用混水器已有一些解(产品), 可立即确定应当选择用节流孔板或节流阀来配置。作为物理效应, 还可以考虑其他效应, 例如, 利用外来能量在热交换器上加热和冷却等等, 但成本比较昂贵而且开发时间较长。



符号说明:  $\dot{V}$ —体积流量  $p$ —压力  $\delta$ —温度  
下标:  $k$ —冷  $w$ —热  $m$ —混合的  $\sigma$ —环境

图 36.1-34 根据图 36.1-33 的要求表构成问题和总功能

接着汇编了关于节流孔板流量和关于同类物质流量混和的物理关系, 见图 36.1-35, 采用同样的物理效应——节流阀或者节流孔板来改变温度和体积流量。在改变混合水流量  $\dot{V}_m$  时必须使体积流量随着水量信号的调位  $S$  作线性的并且同义的改变。与此同时, 温度  $\delta_m$  必须保持不变, 亦即比值  $\dot{V}_k/\dot{V}_w$  必须保持为常数, 并且不允许与信号的调定位置  $S$  相关。

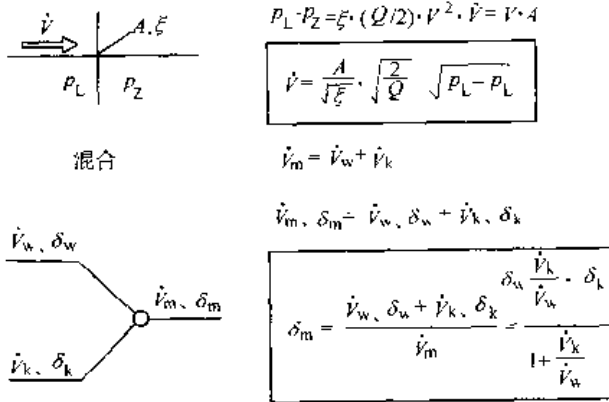


图 36.1-35 关于节流孔板流量和关于同类物质流量混合温度的物理关系

在改变混合水温度  $\delta_m$  时, 体积流量  $\dot{V}_m$  应不改变, 亦即和值  $\dot{V}_k + \dot{V}_w = \dot{V}_m$  必须保持为常数。为此需改变的体积流量  $\dot{V}_k$  和  $\dot{V}_w$ , 必须随标志混合温度的信号调位  $S_\delta$  作线性的变化, 并且这两个体积流量必须作相互反义的变化。

工作步骤3: 建立功能结构 从认识出的分功能, 建立一个初步的功能结构:

- 关断 — 配置 — 混和
- 调节流量
- 调节混和温度

依据已知的物理效应, 目前特别要在几何特征标志方面开发和变异功能结构, 以识别出最好的系统状态: 图 36.1-36~38。从中选出图 36.1-38 的功能结构, 因其系统状态最佳。

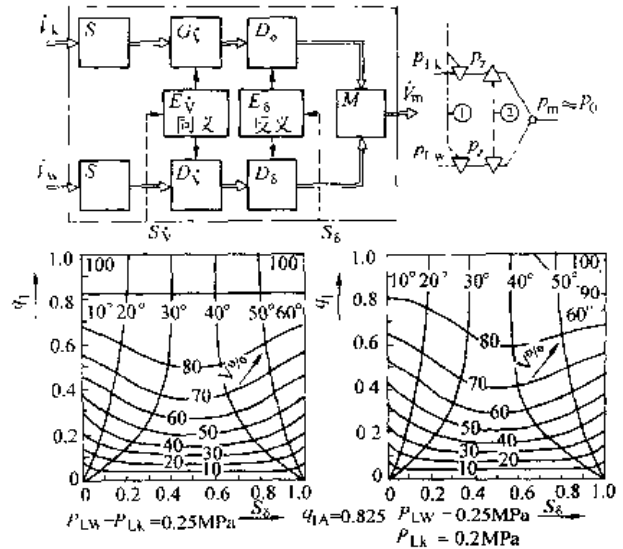


图 36.1-36 根据图 36.1-34 作出的单手把混水器功能结构

在混合前分开两处进行配置①和调温②。在线图中, 以相对调节温度 ( $S_\delta$ ) 和相对调节流量 ( $q_v$ ) 相应于  $S_v$  为转移, 画出等温线和等相对流量线。由于 1 和 2 中的节流孔板处的压力相互影响, 除了设计点 ( $q_{vA} = 0.825$ ) 以外, 温度特性和流量特性都不是线性的, 在流量小时不能用。在冷、热水支路有压力差 (此处为 0.05MPa) 时, 等值线移位, 此时, 即使在设计点上, 两种调节亦不再相互独立 (右面的线圈)。

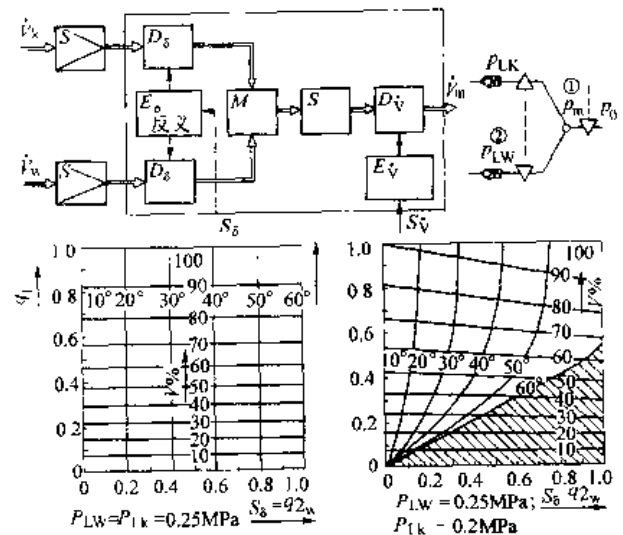


图 36.1-37 根据图 36.1-34 作出的功能结构

在混合前调节温度, 混合后配置。在两个进水管中有相同的前置压力时, 因为在温度—配置孔板处总有恒定的压力差, 所以流量调节和温度调节相互独立, 性态是线性的。但当

前置压力不同时,特性不再是线性的,特别在流量小时强烈移位,因为混和室接近于较小的前置压力。如果超过它的话,不论缸度调节如何,只流冷水或(此处为)热水。

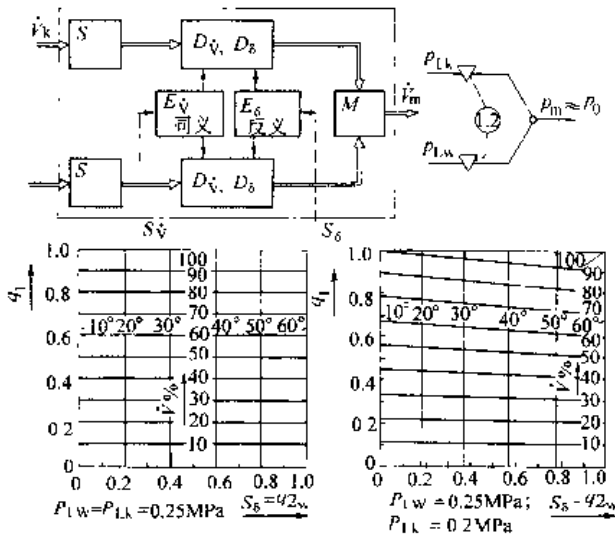


图 36.1-38 根据图 36.1-34 作出的功能结构

温度调节和流量调节在同一些孔板上相互独立地进行,然后再混合。线性的温度特性和流量特性,即使有不同的前置压力,亦无严重改变

工作步骤 4: 按求解的原理实现分功能 由图 36.1-38 的功能结构中的分功能“流量和温度的配定”产生了一个任务:“用一个运动使两个截面作同义改变,同时或者接着用第二个独立运动使其作反义改变”。对这个任务进行了一次智暴,作为初步求解。过程和结果如图 36.1-39 所示。

分析智暴结果,智暴中建议的解,检验了  $V$  和  $\theta$  调节是否有相互独立性。从关于可认识到的运动偶合的分析中,显露出下列作用原理:

1. 对于  $\dot{V}$  和  $\delta$  用两个切于配合面的运动所构成的解

只有用平行于两种调节运动的两对边框来作为节流截面边界线时,才保证  $\dot{V}$  和  $\delta$  调节的相互独立性。这就决定了必须使两个运动互成一角度并且是直线运动。因此每一节流处有两两平行的四条直线边框(图 36.1-40),如此避免了在一种调节运动的同时发生另一运动方向上的变化。

边界框线的分配:构成节流截面的每一构件,必须至少要有两条互成直角而沿运动方向的边线。

在  $\dot{V}$  调节时,两个节流面必须同时朝零变化。

在  $\delta$  调节时,一个面要向零变化,同时另一个面必须达到相应的最大值  $\dot{V}_{max}$ 。

由此可知,在  $\dot{V}$  调节时,两个节流处的边界框

— 圆柱形 — 臂  
轴向运动 =  $\delta$   
旋转运动 =  $\dot{V}$

— 称梁  
— 称梁的逆变  
— 圆柱管的逆变

— 两个面

称梁及塞了  
互作相反运动的阀  
通过剪形件、齿条、  
心轴来制动

— 楔形滑阀 — 平板滑阀  
— 平板滑阀的逆变

(如同称梁)

— 通过凸轮使导管中的球动作  
(圆锥凸轮)

旋转并带轴向运动  
(为了孔板特性,用锐缘)

— 两个面

— 喷射泵(不进一步开发) — 节流阀

两个节流阀  
— 三通混水件  
— 削斜的圆柱形

— 旋转和偏摆

— 升关杆  
球

中心孔  
偏心孔

— 两根软管(用蛋形件  
或楔来挤压)

— 楔在两个孔口间活动  
— 膜片

原则上有两种可能性

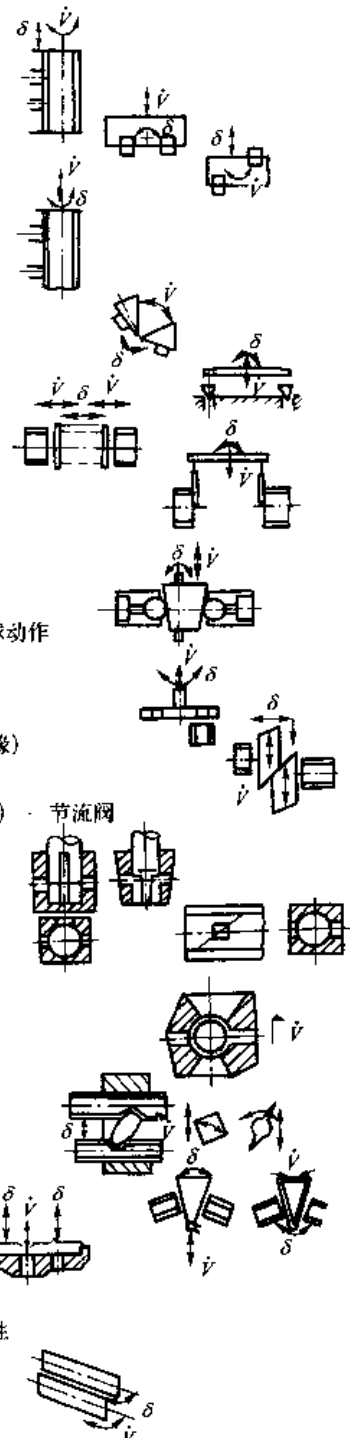


图 36.1-39 为任务“用一个运动使两个截面作同义改变,同时或接着用第二个独立运动使其作反义改变”寻求作用原理的一次智暴会议结构

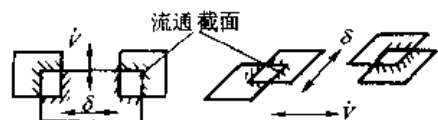


图 36.1-40 节流外的运动和边框线  
线对必须同义地合拢或分开。在  $\delta$  调节时,两处的边界

框线对必须反义地运动,即:一个节流处合拢的同时,另一个节流处分开。

配合面可以是平面的、圆柱的或球面形的曲面。

这种解可以用一个元件作为节流机构,结构上看来是简单的。

2. 对于  $\dot{V}$  和  $\delta$  用两个正交于配合面的运动所构成的解

——此处是指能实现从配合面脱离开的一切运动,但只可能有 3 种运动是垂直于配合面的。

——  $\dot{V}$  和  $\delta$  调节的相互独立性,只有用附加的控制元件(耦合机构)才可能实现。

——结构上的费用看来较大

3. 对于  $\dot{V}$  和  $\delta$  用一个切于配合面的运动所构成的解

——为了保证  $\dot{V}$  和  $\delta$  调节的独立性,此种解亦需有附加的元件以作耦合之用。

——其解在构成上相应于 2 中的。两者的区别只在于配合面形状和由此而定出的运动。

4. 对于  $\dot{V}$  用一个正交于配合面的运动而对  $\delta$  用一个切于配合面的运动所构成的解

——这种解(也不用耦合机构)不满足关于  $\dot{V}$  和  $\delta$  调节的独立性的要求,功能未获保证。

第 1 组的解“对于  $\dot{V}$  和  $\delta$  用两个切于配合面的运动”有明确的性态,同时看来花费较少,因此只对这一种解继续开发下去。不必再在形式上进行一次选择过程。相反,要探讨适当的作用体和运动种类。分析得出图 36.1-41 的编排依据,删去(-)那些看来较不合适的特征标志后,获得图 36.1-42 的编排表式中所所示的、采用不同作用体和作用运动而成的、可能的作用原理。

标称	附件依据	有关的参数
行的	作用体形状	板 楔 (-) 圆柱 圆锥 (-) 球 特殊形状的弹性体 (-)
列的	运动的耦合	直接(准一元件) 间接(机构) (-)
	运动	$\left. \begin{matrix} \dot{V} \\ \delta \end{matrix} \right\}$ 在一个元件上 $\left. \begin{matrix} \dot{V} \\ \delta \end{matrix} \right\}$ 在不同元件上 (-)
	$\dot{V}$ 和 $\delta$ 的运动方向	垂直于配合面 ( $\uparrow$ ) (-) 切于配合面 ( $\rightarrow$ )
	$\dot{V}$ 和 $\delta$ 的运动种类	移动 旋转

图 36.1-41 编排单手把混水器作用原理用的编排依据和有关的参数

运动种类 \ 作用体形状	移/移	移/转	转/转
	1	2	3
平板	A		
圆柱	B		
锥	C		
球	D		

图 36.1-42 单手把混水器解的编排表  
运动方向与配合面相切

对  $\dot{V}$  和  $\delta$  用两个不相耦合面互成直角的运动

工作步骤 5: 选择合适的作用原理 所有作用原理均满足要求表中各项必达要求并预期费用可容许。因此将所有三种作用原理均具体化为原理解。

工作步骤 6: 具体化为原理解的变型

在亦同时研究可能的调节或操作机构(此处未叙述)的情况下,将作用原理具体化为各种可能的结构变型,直至使其成为可进行评判的方案变型,见图 36.1-13~46。

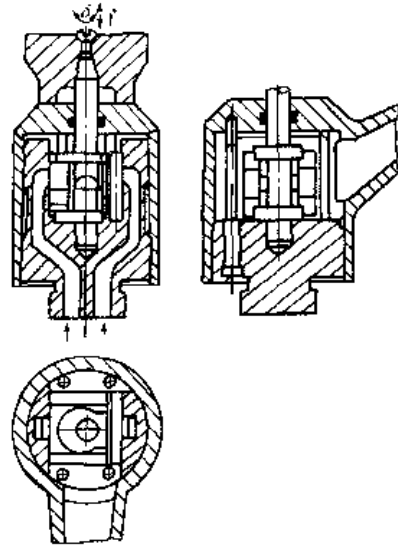


图 36.1-43 单手把混水器解决变型 A:  
“平面解,用偏心片和提转手把”

工作步骤 7: 评价原理性解决变型 按准则 VD12225,借助一张评价表进行评价。再考查评价不可靠性和弱点,见图 36.1-47。

评价的结果,由于图 36.1-44 所示的解 B 有较均衡的价值剖面图和可以看到的改进可能性,获得优选。至于图 36.1-46 所示的球面解 D,则只有在进一步研究并补全了制造和装配方面的信息缺陷,从而作出较肯定的判断后,才值得给予重视。

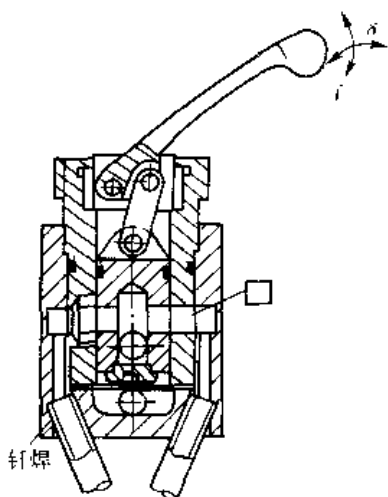


图 36.1-44 单手把混水器解决变型 B:  
“圆柱面解, 用操纵杆”

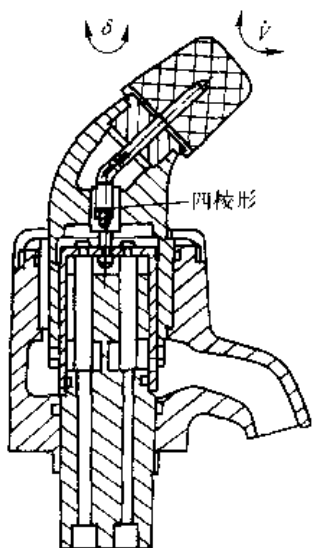


图 36.1-45 单手把混水器解决变型 C:  
“圆柱面解, 用端面阻流”和附加的密封

工作步骤 8: 方案设计阶段结果 对解 B 进行按比例的技术设计, 在占用空间、便于清洁和减少零件方面改进操纵杆。对于解 D, 改进信息状况, 再提交最终判断。

## 6.2 冲击试验台方案设计及详细设计案例分析

### 工作步骤 1: 澄清任务书及编制要求表

任务书的内容是为某研究所开发一种试验机, 这种试验机应能以规定的扭矩冲击性地给轴一载 键联接加载。在制订要求表以前, 必须澄清下述问题:

- 如何理解冲击性负荷?
- 在回转机械中实际上出现什么样的扭矩冲击?

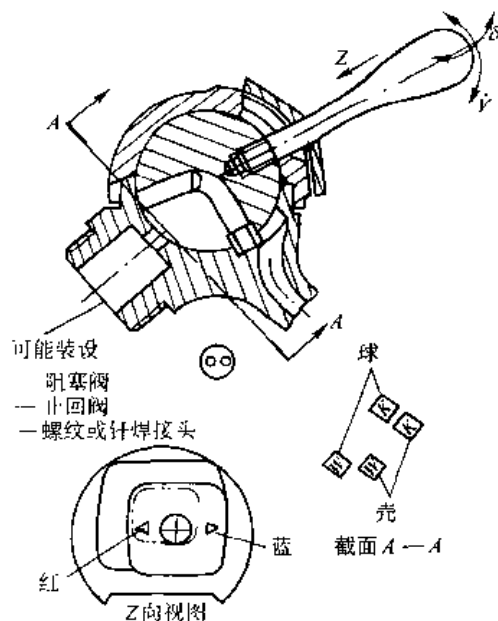


图 36.1-46 单手把混水器解决变型 D:  
“球面解”

在键联接中采取哪种应力测量方法是可能的和合适的?

第一个问题归结为物理基础中关于冲击的概念和两个或更多质量相撞的可能性, 第三个问题由晚些时候在基本研究中给了说明, 然而对于第二个问题则必须进行大量的调查方才可以给予答复。由于试验台应能依照图 36.1-48 对确定的扭矩冲击在高度和梯度上进行调整, 因而实际上的扭矩变化过程首先应在其最大的冲击梯度  $dT/dt$  方面进行分析。为此, 对车床、起重机械、农业机械和轧钢机在非稳定运行状态下的扭矩变化过程进行了观察, 求得最大梯度为  $dT/dt = 125 \cdot 10^3 \text{ Nm/s}$ , 以此作为试验台计算的基础。

在预先澄清了上述问题之后, 就有可能制订一个要求表, 见图 36.1-49。

### 工作步骤 2: 为认识本质的问题而进行抽象化

对要求表逐步抽象化, 以便认识普遍有效的、求解的本质性问题。

第一和第二抽象步骤: 抛弃愿望以及与功能和重要条件无关的要求。

第三抽象步骤: 将定量的说明简化为本质的定性说明。

— 第四抽象步骤: 有意义地扩展已获得知识。

— 第五抽象步骤: 对问题解作不带偏见的描述。

在表 36.1-34 中给出了这些抽象步骤的结果。

### 工作步骤 3: 建立功能结构

功能结构的建立从阐述总功能时即已开始, 总功

		评 价 表 用于单手柄混水器										张: 1/页: 1			
按导则的主要特征标志编排		P: 已有的变量 (P); 改进后可能的		A		B		C		D		E		F	
序号	评价准则	g	P	(P)	P	(P)	P	(P)	P	(P)	P	(P)	P	(P)	
功能	1 关断后不滴水的可靠性	1	1		3		3	4	1						
作用原理	2 可靠而有重复性的调节 (不受碱的影响磨损件少)	1	2		3		2	3	3						
结构形状	3 占地位少 (改装后同样)	1	3↑		2		2		4						
制造	4 件数少	1	1		2		1		1						
	5 制造简单	1	1		3		2		1?	4					
装配安装	6 安装、装配容易	1	2		3		2		2↑	3					
									B						
使用	7 操作舒适 (操作意义明确, 调节细腻, 操作力小)	1	1		3		4		2						
	8 照料容易 (容易洁净)	1	4↓		2		3		2						
维修	9 维护简单 (普通工具, 不拆卸下装置附件)	1	1		3		2		1?	3					
							S		B						
	10														
	11														
	12														
	13														
	14														
?	评价不可靠	$P_{max}=4$		$\Sigma$	16		24	(26)	21	(23)	20	(26)			
↑	趋势较好	$W_g$			0.45		0.67		0.58		0.56				
↓	趋势较差	名次			4		1	(1)	2	(3)	3	(2)			
注或理由 (B), 弱点 (S), 对变型或准则的改进 (V)															
C1	装配用橡胶密封														
B4	简化杠杆机构														
D6	装配时球的位置不固定														
D9															
B8	以 B4 改进														
D9	杆的装固不便于装配														
决策	按比例进一步跟踪研究解 B, 并改进操作元件 研究解 D 的制造可能性, 样品 (2 个月内)														
日期: 10.11		制表人:													

图 36.1-47 单手柄混水器, 评价原理解变型 A, B, C, D

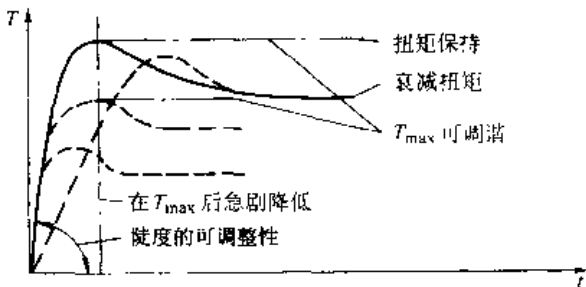


图 36.1-48 扭矩冲击的调整量: 陡度、大小及扭矩持续时间

能则直接来自问题的说明, 见图 36.1-50。满足此复杂的总功能的重要分功能首先与能量流有关, 对于测量的问题, 则与信号流有关:

- 将输入能量转变为负荷值 (扭矩)。
- 将输入能量转变为实现控制功能的辅助能量。
- 贮存能量。
- 控制负荷能量或负荷值大小。
- 改变负荷值。
- 传递负荷能量。

		冲击试验台要求表		第 1 篇	第 1 页
修改	必/愿	要 求		负责人	
10.1	必	几何			
	必	试验联接应空间固定			
		试验轴直径: $\leq 100\text{mm}$			
		(键的尺寸根据 DIN6885 取)			
	必	轮毂负荷传递可沿轴向变化			
		运动学			
	必	负荷在轴静止时加入			
	必	只受单向负荷 (脉动负荷)			
	愿	加载方向可选择			
	愿	扭矩传递可选择由轮毂传至轴、或由轴传至轮毂			
	力				
必	轴-毂-键-联接只受到纯扭转负荷				
	(即无横向力和弯矩影响)				
必	最大扭矩至少保持 1 秒钟				
必	施加负荷的频繁度 (负荷频率): 低				
	(理由: 测量原理)				
愿	轴-毂-键-联接系统的振动在相当大程度上加以排除				
必	最大扭矩可调整到相当于 $\phi 100\text{mm}$ 轴的承受能力 $15000\text{Nm}$ (5 级)				
必	在扭矩最大值后能够实现扭矩急剧降落				
必	扭矩陡度 $ \frac{dT}{dt} $ 可调整 (5 级), 最大值 $\frac{dT}{dt} = 1.25 \times 10^4 \text{N} \cdot \text{ms}$				
必	扭矩变化过程尽可能好地重复实现				
愿	应能达到塑性变形, 必要时能使联接产生破坏				
	能量				
必	功率消耗 $\leq 5\text{kW}/380\text{V}$				
	(极限接载功率)				
	材料				
愿	轴和轮毂材料: Ck45				
		替代 $\times$ 年 $\times$ 月 $\times$ 日 第 $\times$ 级			

图 36.1 49 冲击试验台要求表 (之一)

		冲击试验台要求表		第 1 篇	第 1 页
修改	必/愿	要 求		负责人	
		信号			
	必	测量值: 联接试验前后的扭矩			
	必	沿着键长方向的表面压力			
	愿	测量值可记录			
		测量位置容易达到			
		安全性和人机工程			
	愿	试验台操作尽可能简单 (即试验台换装试件迅速、简单)			
	愿	试验台工作原理对环境无害			
		(低噪声、低污染、低振动……)			
		制造和检验			
	必	所有零件为单件生产			
	必	轴-毂-联接质量符合 DIN6885 (有关规定部分), 其他符合减速器、电动机等轴毂标准: DIN748, 第 2 与第 3 篇			
	愿	试验台按照自己的车间的条件加工			
	愿	尽可能采用外部件和标准件			
		装配和运输			
	愿	试验台: 尺寸小			
		重量轻			
	愿	不带地基			
		使用和维护			
	愿	磨损件少而且简单			
	愿	尽可能无需看管			
		价格			
		造成成本 $\leq 20000$ 马克 (见科研申请)			
		期级			
	必	方案设计阶段结束时期 1973 年 7 月			
		方案设计阶段结束时期 1973 年 7 月 20 日			
		替代 $\times$ 年 $\times$ 月 $\times$ 日 第 $\times$ 级			

图 36.1 49 冲击试验台要求表 (之二)



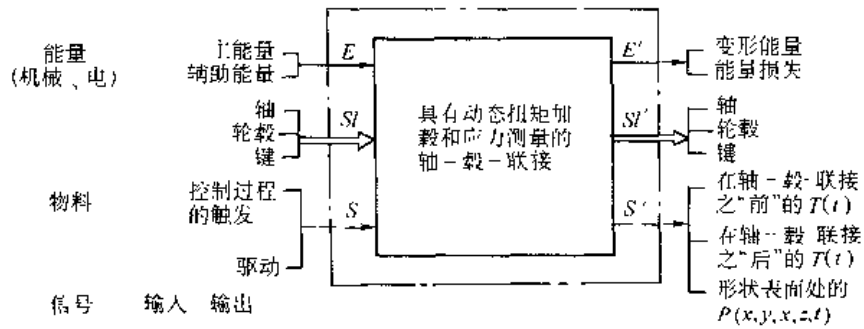


图 36.1-50 冲击试验台的总功能

表 36.1-34 根据图 36.1-49 中要求表所作的抽象和问题阐明

步骤 1 和 2 的结果:

- 试验轴直径  $\leq 100\text{mm}$
- 轮毂传力沿轴向变化
- 负荷在轴静止时加入
- 试验联接受到可调整的纯扭矩载荷, 最大值达  $15000\text{N} \cdot \text{m}$ , 最大扭矩至少保持 3s
- 扭矩必须能够冲击式地降落
- 最大可能的扭矩梯度为,  $dT/dt = 1.25 \times 10^5 \text{N} \cdot \text{m/s}$
- 扭矩变化过程重复实现
- $T$  前,  $T$  后和  $p$  的测量值可以记录下来

步骤 3 的结果:

- 轴-毂-键联接的扭矩负荷在高度、梯度、保持时间和扭矩降落方面均可调整
- 扭矩及应力试验应在轴静止时进行

步骤 4 的结果:

- 进行构件试验时动态扭矩可调整
- 能够测量输入负荷和构件应力

步骤 5 的结果:

“以动态扭矩给轴-毂-键联接加载, 并且测量其应力。”

- 将负荷作用在试件 (作用表面) 上。
- 测量负荷。
- 测量构件应力。

通过对这些分功能进行组合, 不同顺序地排列、补充或剔除, 可以逐步地形成多个功能结构变型, 图 36.1-51 按照其生成顺序表示了这些功能结构变型。由于在设计任务书中对测量的解没有限定, 因而仅仅在能量流和物料流方面发展功能结构。下面将要进行的是寻找功能结构变型 4 的解, 因为这个变型包含了同样值得探索的变型 5 的分功能。在求解以前, 将单个的分功能进行细分, 例如将“能量转变”分成“扭矩转变为力”, “力的传递”, “力反转变为扭矩”, 看来是不合适的。

工作步骤 4: 寻找实现分功能的作用原理

为了寻找作用原理, 首先应考虑下述工作:

- 传统的辅助手段:
  - 文献调查, 分析已有的万能试验机。
- 强调直觉的方法:
  - 智暴法。
- 强调推理的方法:
  - 借助于编排表、能量类型、作用运动和作用表面的

特殊变型进行系统化搜索, 利用力变化原理的目录。

为了对找到的作用原理进行综合, 做出一编排表, 见图 36.1-52。这里仅仅从大量内容中, 记入最重要的分功能和作用原理。从开始时即筛除或者从表上划掉那些用不着的作用原理, 特别是对机械的控制原理, 其他的作用原理可通过有目的地定出作用表面和作用运动的变型而得出。对于本例来说, 不论对于“能量转变”分功能, 或者对于其他分功能, 能量的种类是对作用原理进行编排的主要出发点。

对于分功能“能量释放”(接通), 作为作用原理可考虑不同形式的可控离合器及棘轮传动装置, 对于分功能“加载”(将负荷传至作用位置), 可考虑带有不同的轴-毂-联接以及刚性联轴器的轴, 对于并不限定解的分功能“测量”, 则首先考虑与相应的电子测量放大器相连的电阻应变片、电感传感器和电容传感器。

工作步骤 5: 将作用原理组合成实现总功能的作用结构

现在必须对利用前述步骤找到的分解组合成总解, 在本例中借助于同一个编排表来做到这点。从作用原理范围中, 通过将某一分功能的解与邻近分功能的解进行组合, 可以导出不同的作用结构。

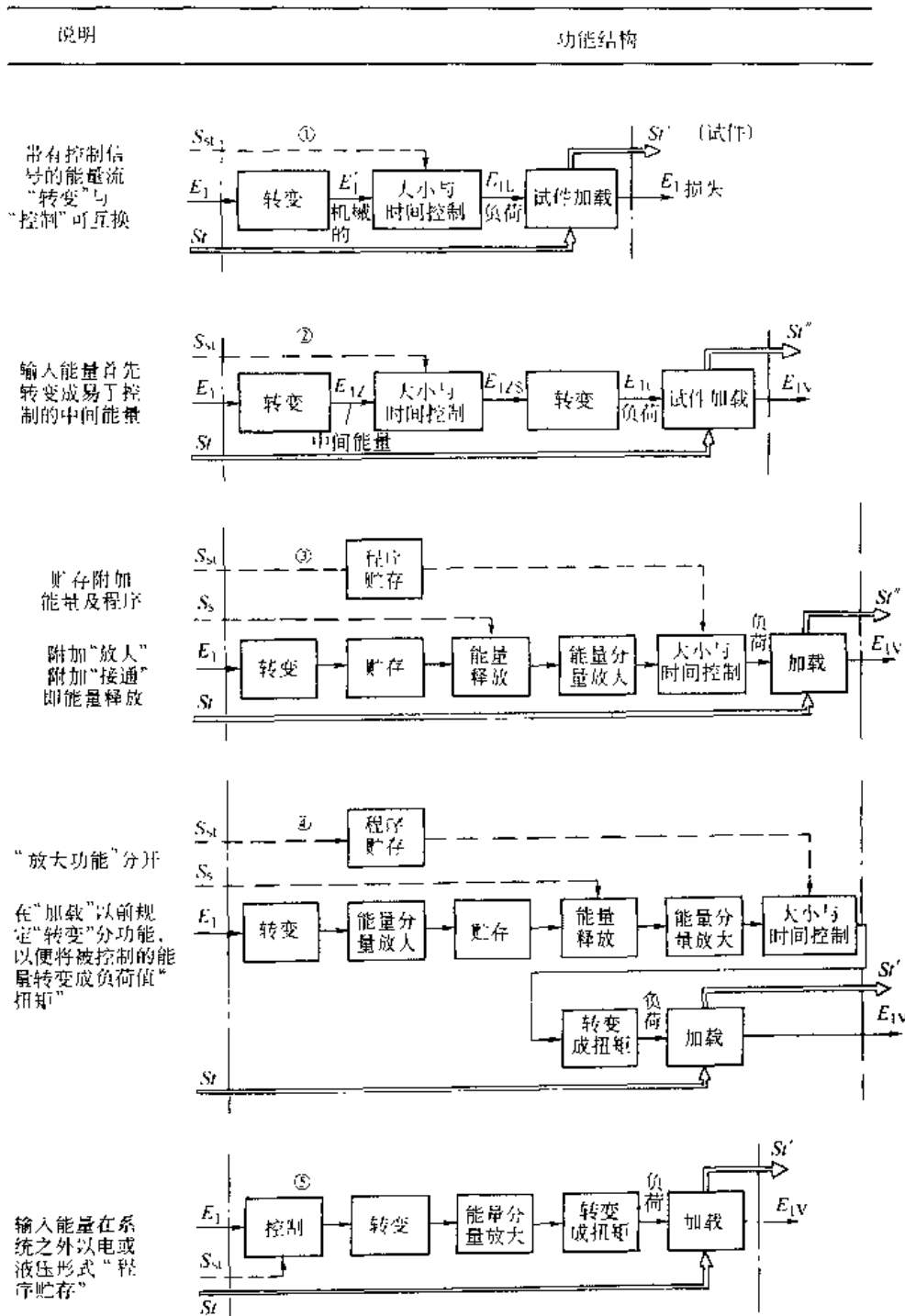


图 36.1-51 逐步建立的功能结构变型

此时，功能结构变型 4 与 5 可用作进行这种结合的基础，其中分功能的排列顺序还可以部分地改变。在组合的过程中，要通过讨论非系统地考虑其相容性和技术上实现的可能。在这里，建立一个相容性矩阵可能是很有帮助的。图 36.1-53 表示了可能实现的功能结构的范围。

**工作步骤 6：选择合适的变型**

当已有了大量变型方案时，在进一步具体化以前

应进行预先选择，使得大量设计草图及专门计算的消耗仅仅花费在看来值得进行下去的作用结构上。图 36.1-54 示出了这样的选择。在首先找到的七种作用结构中，经过选择后仅有四种将继续进行下去。

**工作步骤 7：解的变型的具体化**

为了能作出最有利的解的变型的可靠决策，必须使得选出的作用结构成为有可能作出判断的形式。首先有必要对已有的作用结构作出原理图，见图 36.1-

作用原理		1	2	3	4	5	6	7	8	9
1	电的 ↓ 机械的	不同型式电动机	直流马达	电致伸缩	磁致伸缩	压电晶体	电容	电磁铁		
		液体静压容积式驱动装置 (泵或马达)	液体动压原理 (泵或汽轮机)	电渗电泳	电致伸缩	磁致伸缩	电容	电磁铁		
2	电的 ↓ 液压的	液体静压容积式驱动装置 (泵或马达)	液体动压原理 (泵或汽轮机)	磁流体效应	电渗电泳	压电晶体	电容	电磁铁		
		螺旋传动	齿轮传动	凸轮机构	曲柄机构	联合传动	突然固定	杠杆	拉绳传动	
3	机械的 ↓ 机械的	螺旋传动	齿轮传动	凸轮机构	曲柄机构	联合传动	突然固定	杠杆	拉绳传动	
		活塞	螺旋泵及马达	齿轮泵及马达	叶片泵及马达	轴向柱塞泵及马达	轴向柱塞泵及马达	液体动压原理	浮力效应	
4	机械的 ↓ 液压的	活塞	螺旋泵及马达	齿轮泵及马达	叶片泵及马达	轴向柱塞泵及马达	轴向柱塞泵及马达	液体动压原理	浮力效应	
		飞轮 (回转)	惯性质量 (平移)	位能	变形	电池	电容 (电场)	液力储能器 a) 加压贮能器 b) 活塞贮能器 c) 膜片贮能器 (压力能)	液体贮石 (位能)	
5	能量贮存 $E(t) \rightarrow E(t+\Delta t)$	飞轮 (回转)	惯性质量 (平移)	位能	变形	电池	电容 (电场)	液力储能器 a) 加压贮能器 b) 活塞贮能器 c) 膜片贮能器 (压力能)	液体贮石 (位能)	
		凸轮机构, 按作用面与作用运动变化	滚动连杆	行星齿轮传动	可控轴动器	欧姆或感应电阻	可控硅	可控流量阀	可控马达与泵	
6	控制能量 大小与时间 $E_f(S)$	凸轮机构, 按作用面与作用运动变化	滚动连杆	行星齿轮传动	可控轴动器	欧姆或感应电阻	可控硅	可控流量阀	可控马达与泵	
		楔	肘杆	连杆	压力放大					
7	能量分量变化	楔	肘杆	连杆	压力放大					

图 36.1-52 用于冲击试验台的编排表 (部分)

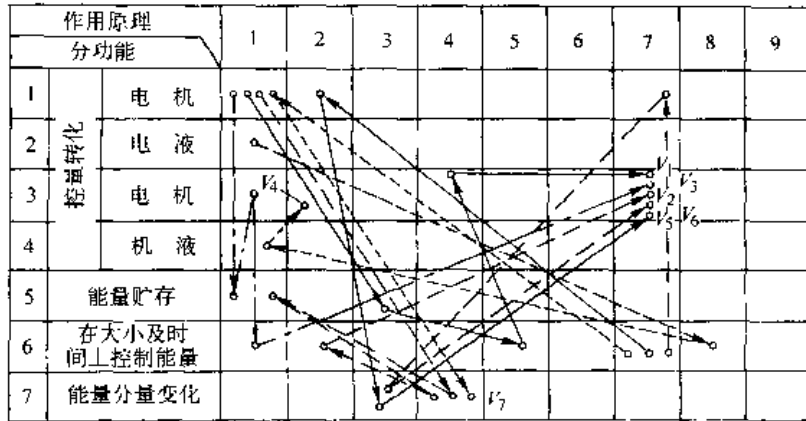


图 36-1-53 按照图 36-1-52 利用作用原理作出的七种作用结构联接图

- 变型 1: 1. 1-5. 3-6. 5-3. 4 3. 7
- 变型 2: 1. 1-7. 4-5. 1-7. 4-6. 2-3. 7
- 变型 3: 1. 1-5. 1-3. 1-6. 1-3. 7
- 变型 4: 2. 1-5. 8-4. 1-3. 2
- 变型 5: 6. 7-1. 2-7. 3-3. 7
- 变型 6: 6. 7-1. 7-7. 3-3. 7
- 变型 7: 6. 7-1. 1-7. 4

		冲击试验台选择表							第 1 页 共 1 页	
		解的变型 (L <sub>V</sub> ) 按选择准则进行判断:							决策	
		(-) 是 (-) 否 (?) 缺乏情报 (!) 检查要求表							(+ 继续跟踪此解 (-) 删除此解 (?) 收集情报(重新判断此解) (!) 检查修改要求表的必要	
		相容性解								
		满足要求表的要求								
		原则上可实现								
		价格上允许								
		满足直接安全技术								
		在本范围内优先								
		知识状态充分								
		需要! 提示、论据)							决策	
L <sub>V</sub>		A	B	C	D	E	F	G		
V <sub>1</sub>	1	+	2	+	+	2			可控制动计算成问题	+
V <sub>2</sub>	2	+	1	1	-	+	-	1		-
V <sub>3</sub>	3	+	-	+	+	-	-	1		+
V <sub>4</sub>	4	+	-	+	2	+	-	-	大功率液压系统尚未应用过	+
V <sub>5</sub>	5	-	2	-		+	-	-	对直线马达没有经验	-
V <sub>6</sub>	6	+	2	+	2	-	-	2	对磁体功率要求过高	-
V <sub>7</sub>	7	+	2	-	-	+	-	2	对晶闸管控制没有经验	-
	8									
	9									
	10									
	11									
	12									
	13									
	14									
	15									
	16									
	17									
	18									
	19									
	20									
日期: 7.15		制订人:								

图 36-1-54 作用结构的选择

35 至图 36.1-58。为了判断某一解的功能效率, 仅作线条草图往往不够。此时, 进行有目的的计算或者模型研究是有用的。对于本设计任务, 作为此进程的举例, 对解的变型  $V_2$  中控制扭矩冲击的凸轮机构及所需的飞轮惯性矩 (能量贮存器) 进行了估算。

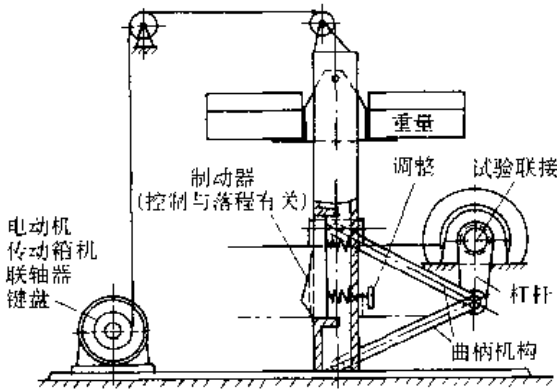


图 36.1-55 解的变型  $V_1$

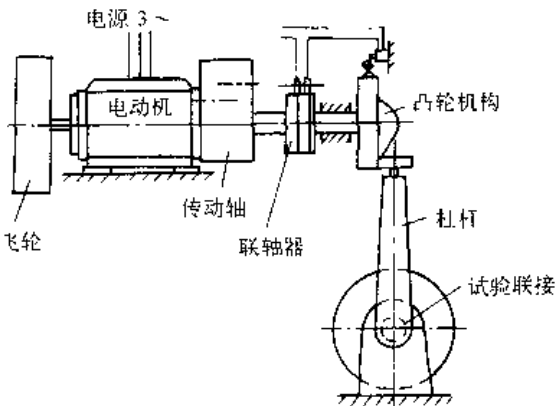


图 36.1-56 解的变型  $V_2$

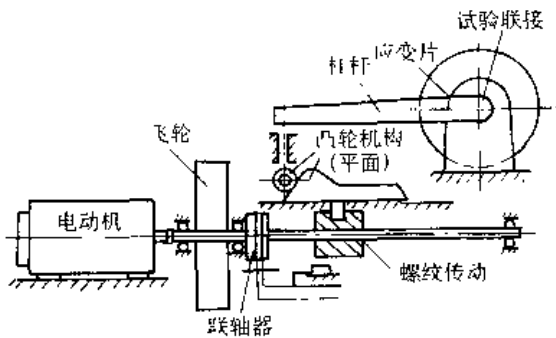


图 36.1-57 解的变型  $V_3$

图 36.1-59 中设计的凸轮, 能否获得  $dT/dt = 1.25 \times 10^3 \text{ N} \cdot \text{m/s}$  的要求的冲击梯度和  $T_{\max} = 1.5 \times 10^3 \text{ Nm}$  的最大扭矩?

计算步骤:

— 在要求的冲击梯度下达到最大扭矩所需要的时间:

$$\text{时间: } \Delta t = \frac{1.5 \cdot 10^3}{1.25 \cdot 10^3} = 0.12 \text{ s}$$

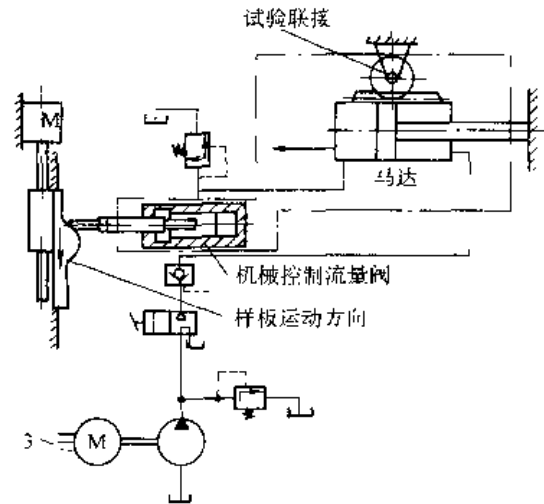


图 36.1-58 解的变型  $V_4$

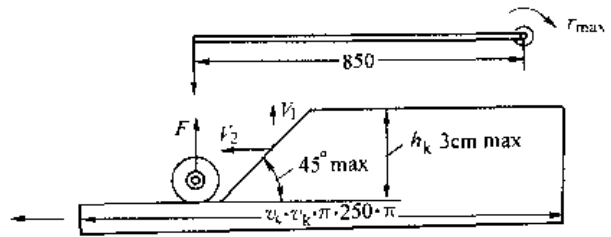


图 36.1-59 凸轮展开曲线

$$\begin{aligned} \text{— 加载杠杆端部受力: } F_{\max} &= T_{\max} / l = \frac{1.5 \cdot 10^3}{0.85} \\ &= 17.6 \cdot 10^3 \text{ N} \end{aligned}$$

作为柔软的弯曲弹簧的加载杠杆应当这样计算, 使得它在所选的凸轮升程  $h = 30 \text{ mm}$  情况下受到力  $F_{\max}$  而弯曲, 此时所受到的弯曲应力不超过其允许值。

— 凸轮的圆周速度:

$$v_s = v_c = \frac{h_k}{\Delta t} = \frac{30}{0.12} = 250 \text{ mm/s}$$

— 凸轮的角速度与转速:

$$\omega = \frac{0.25}{0.125} = 2.0 \text{ s}^{-1}; n_k = \frac{60\omega}{2\pi} = 19 \text{ min}^{-1}$$

— 每圈回转时间:

$$t_u = \frac{2\pi}{\omega} = 3.14 \text{ s}$$

由于接入和脱开凸轮机构的电磁可控离合器的作用时间, 在十分之几秒的范围内, 因而实现这个原理不存在困难。扭矩冲击的高度和梯度可以方便地通过更换凸轮以及改变回转时间而加以改变。

估算飞轮惯性矩的计算步骤如下:

— 估算冲击时所需要的能量及为冲击而贮存的能量, 是假设处于力流中所有零件都产生弹性变形的条件下进行的。

在最大冲击扭矩下所贮存的能量:

$$W_{\max} = \frac{1}{2} F_{\max} \cdot h_k = 260 \text{ Nm} = 260 \text{ Ws}$$

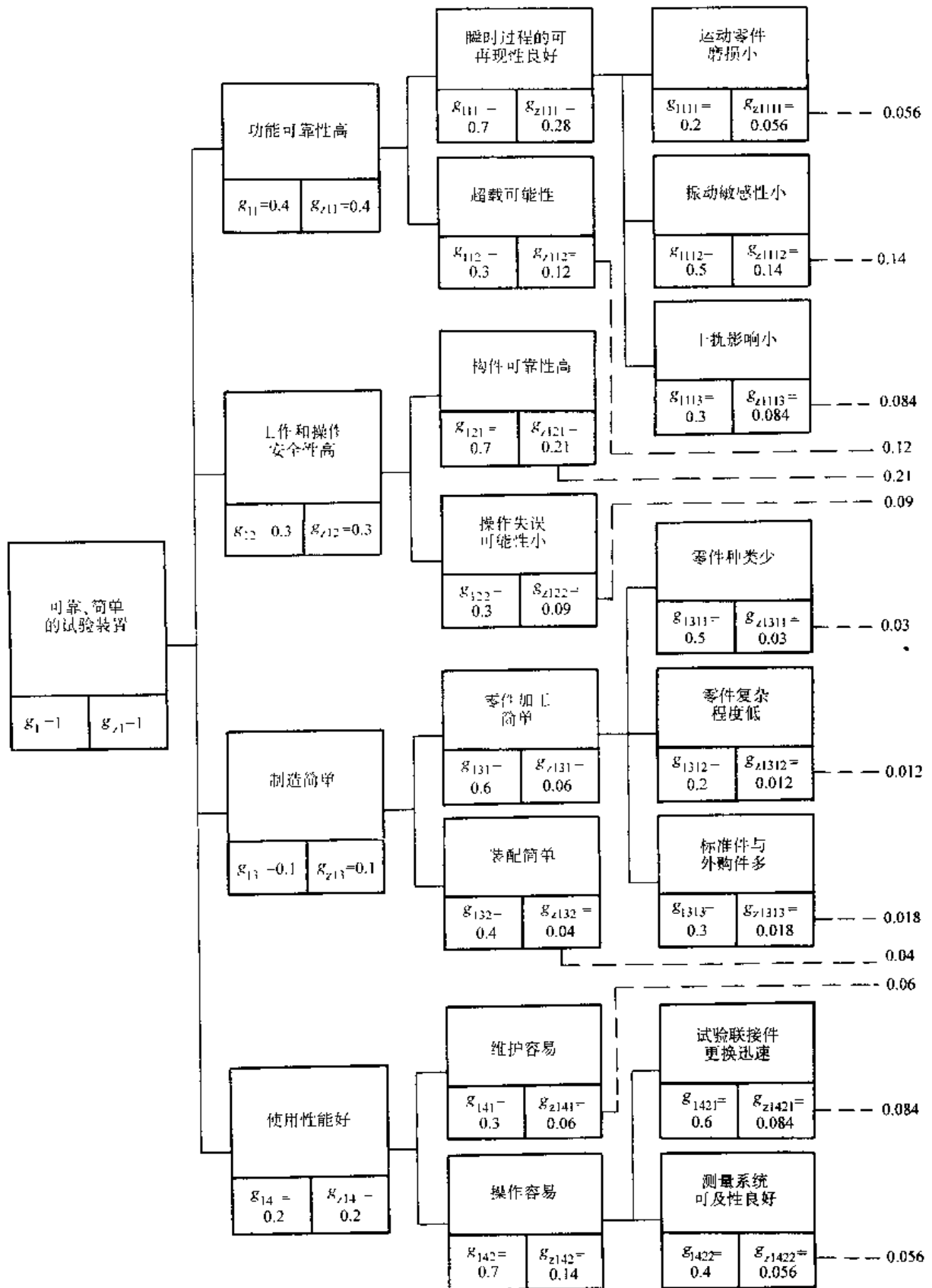


图 36.1-60 冲击试验台的目标系统

序号	评价准则	加权系数	特性值		变型 V <sub>1</sub>			变型 V <sub>2</sub>			变型 V <sub>3</sub>			变型 V <sub>4</sub>		
			单位	特性 c <sub>1</sub>	价 w <sub>01</sub>	加权价 w <sub>01c1</sub>	特性 c <sub>2</sub>	价 w <sub>02</sub>	加权价 w <sub>02c2</sub>	特性 c <sub>3</sub>	价 w <sub>03</sub>	加权价 w <sub>03c3</sub>	特性 c <sub>4</sub>	价 w <sub>04</sub>	加权价 w <sub>04c4</sub>	
1	磨损少	0.056	-	高	3	0.168	0.336	中	6	0.336	0.336	低	6	0.336		
2	振动敏感性低	0.14	s <sup>-1</sup>	410	3	0.420	0.980	2370	7	0.980	2370	<410	2	0.280		
3	干扰影响少	0.084	-	高	2	0.168	0.588	低	7	0.588	0.504	中	4	0.336		
4	超载可能性	0.12	%	5	5	0.600	0.840	10	7	0.840	0.840	20	8	0.960		
5	构件可靠性高	0.21	-	中	4	0.840	1.470	高	7	1.470	1.470	很高	8	1.680		
6	操作失误可能性低	0.09	-	高	3	0.270	0.630	低	7	0.630	0.540	中	4	0.360		
7	零件种类少	0.03	-	中	5	0.150	0.120	中	4	0.120	0.120	低	6	0.180		
8	零件复杂程度低	0.012	-	低	6	0.072	0.084	低	7	0.084	0.060	高	3	0.036		
9	标准件与外购件多	0.018	-	低	2	0.036	0.108	中	6	0.108	0.108	高	8	0.144		
10	装配简单	0.04	-	低	3	0.120	0.200	中	5	0.200	0.200	高	7	0.280		
11	维护容易	0.06	-	中	4	0.240	0.480	低	8	0.480	0.420	高	3	0.180		
12	更换试验联接件迅速	0.064	min	180	4	0.336	0.588	120	7	0.588	0.588	180	4	0.336		
13	测量系统可及性好	0.056	-	好	7	0.392	0.392	好	7	0.392	0.392	中	5	0.280		
		E <sub>g1</sub> = 1.0			G <sub>w01</sub> = 51 W <sub>1</sub> = 0.39	G <sub>w02</sub> = 85 W <sub>2</sub> = 0.65	G <sub>w03</sub> = 6.816 W <sub>3</sub> = 0.68		G <sub>w04</sub> = 68 W <sub>4</sub> = 0.52	G <sub>w01</sub> = 78 W <sub>1</sub> = 0.60	G <sub>w02</sub> = 6.446 W <sub>2</sub> = 0.64		G <sub>w03</sub> = 68 W <sub>3</sub> = 0.52	G <sub>w04</sub> = 5.388 W <sub>4</sub> = 0.54		

图 36.1-61 冲击试验台四种解的变型的评分

这一能量值是在  $\Delta t = 0.12s$  的时间间隔内所需要的。

——飞轮尺寸；

选择最大转速：

$$n_{max} = 1200\text{min}^{-1}; \omega \approx 126\text{s}^{-1}$$

由选出的飞轮尺寸  $D_A = 0.4\text{m}$  和  $B = 0.3\text{m}$  得出飞轮质量  $m = 100\text{kg}$ 。由此惯性矩为：

$$J = \frac{1}{2} m_A r^2 = 2\text{kgm}^2;$$

飞轮贮存的能量：

$$W = \frac{1}{2} J \omega^2 = 159.10^3 \text{Nm} (\text{即 } W_s)$$

——冲击后的转速下降：

$$W_{Rest} = W_s - W_{max} = 15640W_s$$

$$\omega_{Rest} = \sqrt{\frac{2W_{Rest}}{J}} = 125\text{s}^{-1}$$

$n_{Rest} = 1190\text{min}^{-1}$ ，亦即转速下降很少，相应地只需要一个小功率的驱动马达就可以了。

**工作步骤 8：解的变型的评价**

在完成具体化的基础上，可以对解的变型的重要特性进行估计，使得其评价具有结论性。评价将在工作步骤 6 中选出的四个解的变型中进行。

从要求表中的重要愿望出发，首先产生了一系列不同复杂程度的评价准则，最后发展成一个级别严格的顺序（目标系统），以便能够较好地认识，并且编排变型方案的加权因数和特性值，见图 36.1-60。

与评价准则有关的主要变型特性值以给定的数值综合列在一个评价表中，见图 36.1-61。

由此得出，变型  $V_2$  具有最高的总价值和最好的总价值比。与其最接近的是变型  $V_3$ ，为了了解薄弱部位，作出一价值轮廓图，见图 36.1-62。由此人们了解到变型  $V_2$  在主要评价准则方面的平衡状态。变型  $V_2$  以 68% 的加权价值比成为最终进行技术设计的最有利的原理解（方案）。

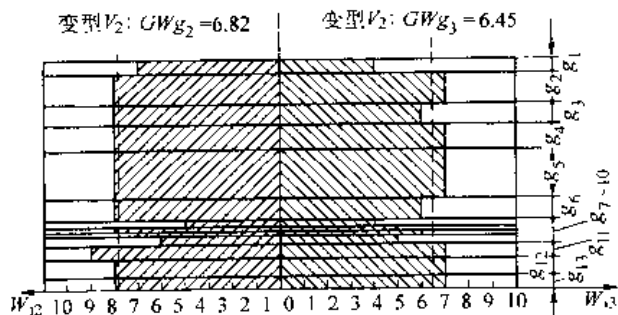


图 36.1-62 用于识别薄弱部位的价值剖面图  
上述工作步骤 1 至 8，描述了方案设计阶段。

**工作步骤 9：识别确定结构的要求，澄清空间条件**  
由要求表（图 36.1-49）可以得知（选出）下述确

定结构的要求：

——确定布置的：

试验联接应空间固定，

只能在轴静止时朝着一个方向加载。

在轮毂端负荷传出可变化。

力矩传入可变化。

本身没有地基。

——确定尺寸的：

试验轴直径  $\leq 100\text{mm}$ 。

可调整扭矩  $T \leq 15000\text{Nm}$ ，5 级（保持时间最少 30）。

可调整扭矩梯度  $dT/dt = 1.25 \times 10^4 \text{Nm/s}$ ，5 级。

功率消耗  $\leq 5\text{kW}$ 。

——确定材料的：

轴和轮毂材料为 Ck45。

——其他要求：

在自己的车间内单件生产。

应用外购件和标准件。

拆卸方便。

对于空间条件没有给出特殊要求。

**工作步骤 10：对确定结构的主功能载体进行结构化**

进行结构化的基础就是功能结构变型 Nr. 4（图 36.1-51）和原理解变型  $V_2$ （图 36.1-56）。

由此可见，作为确定结构的主功能载体，下列功能是很重要的：

——试验联接，在凸轮与试验联接轴之间的传力杠杆。

——凸轮。

其它的主功能载体是：

——电动机。

——飞轮。

——可控联轴器。

——传动箱。

——试验台机架。

**工作步骤 11：确定结构的主功能载体的粗略结构设计**

表 36.1-35 为主功能载体。图 36.1-63 首先表示了三个确定结构的主功能载体的粗略布置图。

试验联接的结构设计可参看 DIN6885，传力杠杆的设计则根据弯曲弹簧的验算而进行，两者在相当程度上没有什么问题，与此相反，凸轮的参数计算和结构设计则需根据提出的要求对其运动关系和动力特性进行深入的分析。由仔细观察可知，在方案设计阶段对凸轮控制功能所作的估算，对于结构设计来说还是不够



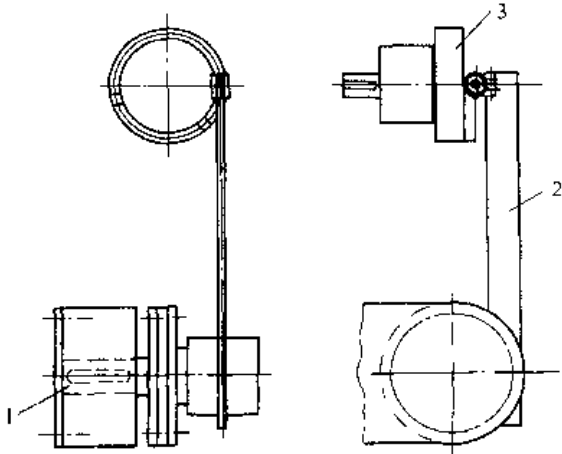


图 36.1-63 三个确定结构的主功能载体的粗略布置图

1—试验联接 2—传力杠杆 3—凸轮

的。更确切地说，在确定主要尺寸以前，要进行下列运动分析：

表 36.1-35 主功能载体

功能	功能载体	特征标志
能量转换， 能量分量放大	电动机	功率 $P_M$ 转速 $n_M$ 起动时间 $t_M$
能量贮存	飞轮	惯性矩 $J_s$ 转速 $n_s$
能量接通	离合器	可传递力矩 $T_{KU}$ 最高转速 $n_{KU}$ 结合时间 $t_{KL}$
能量分量放大	传动箱	功率 $P_G$ 最大输出扭矩 $T_G$ 输出转速 $n_G$ 传动比 $i_G$
大小和时间控制	凸轮	功率 $P_K$ 扭矩 $T_K$ 转速 $n_K$ 直径 $D_K$ 升角 $\alpha_K$ 高度差 $h_K$
将能量转变成扭矩	杠杆	长度 $l_H$ 刚度 $c_H$
试验联接件加载	试验联接	扭矩 $T$ 扭矩梯度 $dT/dt$
承受力和扭矩	机架	—

按照图 36.1-64，有：

试验轴所受扭矩：

$$T = c_H \cdot h_K \cdot l_H$$

扭矩梯度：

$$\frac{dT}{dt} = \pi \cdot D_K \cdot n_K \cdot \text{tg}\alpha_K \cdot c_H \cdot l_H$$

持续时间：

$$l_H = \frac{U_K}{2\pi \cdot D_K \cdot n_K} = \frac{1}{2 \cdot n_K}$$

上述扭矩梯度的关系式，只有在杠杆移动轨迹与凸轮轮廓轨迹相平行的前提下方才有效（图 36.1

65）。由于希望获得较低的摩擦损失而采用滚子在凸轮轮廓上扫描，使得实际的扭矩梯度比计算值低，并且不是常数，所以可以按照图 36.1-66 计算出一个平均梯度。

如果要求表的扭矩梯度  $dT/dt$  是根据图 36.1-66 的平均扭矩来计算，则在计算  $dT/dt$  时，可不再代入整个的圆周速度  $V_x$ ，而仅代入有效圆周速度  $V_x^*$ 。

此时有：

$$v_x^* = K \cdot v_x$$

修正系数  $K$  与下述因素有关：

- 升角  $\alpha_K$ ，
- 滚子直径  $d$ ，
- 凸轮的高度差  $h_K$ 。

修正系数  $K$  可由图 36.1-67 进行推导：

利用 
$$x = \frac{h_K}{\text{tg}\alpha_K}$$

及 
$$x = d/2 \left( \sin\alpha_K - \frac{1 - \cos\alpha_K}{\text{tg}\alpha_K} \right)$$

可得 
$$K = \frac{v_x^*}{v_x} = \frac{x}{x + \Delta x}$$

（上式仅在  $d/2 \cdot (1 - \cos\alpha_K) \leq h_K$  的情况下才准确）

亦即

$$K = \frac{\frac{h_K}{\text{tg}\alpha_K}}{\frac{h_K}{\text{tg}\alpha_K} + d/2 \cdot \left( \sin\alpha_K - \frac{1 - \cos\alpha_K}{\text{tg}\alpha_K} \right)}$$

为了得到  $K$  的数值，取下列各参数估计值：

升角  $\alpha_K = 10^\circ \sim 45^\circ$

滚子直径  $d = 60\text{mm}$

— 凸轮高度差  $h_K = 30\text{mm}$  及  $h_K = 7.5\text{mm}$

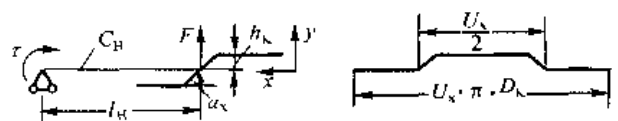


图 36.1-64 凸轮与杠杆的几何关系； $c_H$  杠杆刚度

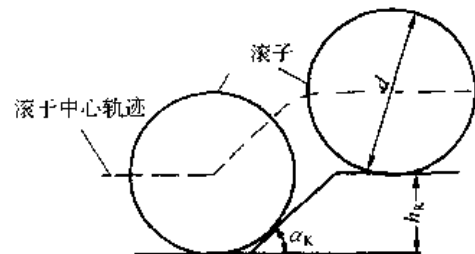


图 36.1-65 凸轮轮廓轨迹和杠杆移动轨迹

由上列公式算出的  $K$  值列于表 36.1-36 中。

将计算扭矩梯度  $dT/dt$  的公式按照凸轮转速  $n_K$  进行移项，并且考虑修正系数  $K$ ，则得：

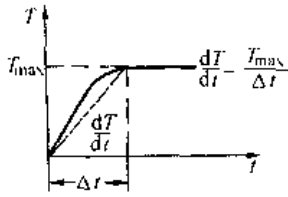


图 36.1-66 扭矩梯度

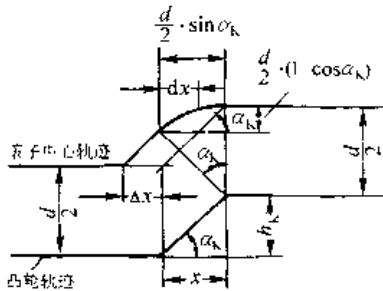


图 36.1-67 修正系数 K 的推导

表 36.1-36 修正系数 K 的参考值

$h_k / \text{mm}$	$\alpha_k$	45	40	30	20	10
7.5	K	0.41	0.45	0.62	0.79	0.94
30	K	0.71	0.76	0.87	0.94	0.98

$$n_k = \frac{dT/dt}{K \cdot \pi \cdot D_k \cdot \text{tg} \alpha_k \cdot C_H \cdot L_H}$$

转速的调整范围由下面的方法确定：

$$R = \frac{n_{kmax}}{n_{kmin}}$$

如果将由解决方案得出的凸轮直径  $D_k$ ，杠杆的刚度  $C_H$  和长度  $L_H$  看成常数，就可以根据前面的公式，利用其余的参数  $dT/dt$ 、 $K$  和  $\alpha_k$  的关系求出凸轮转速  $n_k$  的两个极限值，见表 36.1-37。

表 36.1-37  $n_{kmin}$  及  $n_{kmax}$  的确定

	$\frac{dT}{dt}$	$\alpha_k$	K	$n_k$
最小	20	10	0.98	$116 \cdot C$
最大	125	45	0.41	$305 \cdot C$

表中  $C$  是考虑单位和其他常数值 ( $\pi$ 、 $D_k$ 、 $C_H$  及  $L_H$ ) 的常数。

由此得到转速调节范围为

$$R = \frac{305 \cdot C}{116 \cdot C} = 2.6$$

这意味着：

— 数量和控制时间的功能，不能仅通过凸轮来实现。

— 在原则上保持方案设计的情况下，必须改变功能结构。

— 凸轮必须有一个调速范围约为  $R=2.6$  的变速驱动装置。

图 36.1-68 表示了通过另一个分功能“转速调节”所补充的功能载体变型 (图 36.1-51)。这一分功能可通过一无级调速驱动马达来实现。此外，还可以提出多个变型方案 (4/1~4/3)。由上述关系式进行的凸轮参数计算，得到下列主要特征标志的数值：

杠杆弹簧刚度  $c_H = 700 \text{N/mm}$ ，杠杆长度  $L_H = 850 \text{mm}$ ，凸轮直径  $D_k = 300 \text{mm}$ ，升角范围  $\alpha_k = 45^\circ \sim 10^\circ$ ，常数  $C = 0.107 \text{min}^{-1}$ ，对于要求的扭矩梯度范围 ( $dT/dt_{max} = 125 \times 10^3 \text{Nm/s}$ ， $dT/dt_{min} = 20.103 \text{Nm/s}$ )，调速范围为  $R = 2.6$ ，转速范围为  $n_k = 12.4 \sim 32.6 \text{min}$ 。

对扭矩梯度  $dT/dt$  进行调整的要求，也必须通过利用所选用的数据加以满足。

对于在最大扭矩下要求的保持时间，情况有所不同，得出  $t_H = (1/2) \cdot n_k = 2.4 \sim 0.92 \text{s}$ ，此要求值  $3\text{s}$  要低。经过与用户进行磋商，将此要求减低至  $t_H \geq 1\text{s}$ ，这通过利用凸轮的大半个圆周就可以实现了。

为了能按比例地描述确定结构的主功能载体，还需要事先澄清以下各点：

- 试验联接和凸轮的空间布置是什么样的？
- 对辅助功能载体应同时考虑到什么程度？

根据下述理由，确定试验联接水平布置，从而凸轮应绕一垂直轴回转：

- 试验联接和凸轮的良好可更换性。

+ 考虑装配的合理结构设计。

- 方便测量试验联接。

- 考虑使用的合理结构设计。

- 试验联接的夹紧力传给地基。

+ 短程直接的力传导。

— 试验台为适应其他形式的试样装调简单 (特别是对较长的试样)。

+ 考虑冒险的合理结构设计，对辅助功能载体的必要性进行估计，其占地面积按经验加以确定。

— 由于在凸轮处产生的轴向力  $F_A$  和圆周力  $F_U$ ，有必要给凸轮一种特殊的支承

$$F_A = F_U = \frac{T_{max}}{r_U} = 17.6 \text{kN}$$

— 试验联接与杠杆之间的法兰联接外径，根据刚性回转离合器必须大致取为  $Da = 400 \text{mm}$ 。

然而可以看出，辅助功能载体对于按比例的结构设计仅产生次要的影响。

图 36.1-69 表示了在功能结构变型 4/1 的基础上作出的粗略结构设计，此时转速的调节通过装在可控式离合器后面的调速传动箱实现。图 36.1-69b 表示了功能结构变型 4/2 基础上作出的粗略结构设计，其中调速传动箱装在可控式离合器的前面。变型 4/3 (图

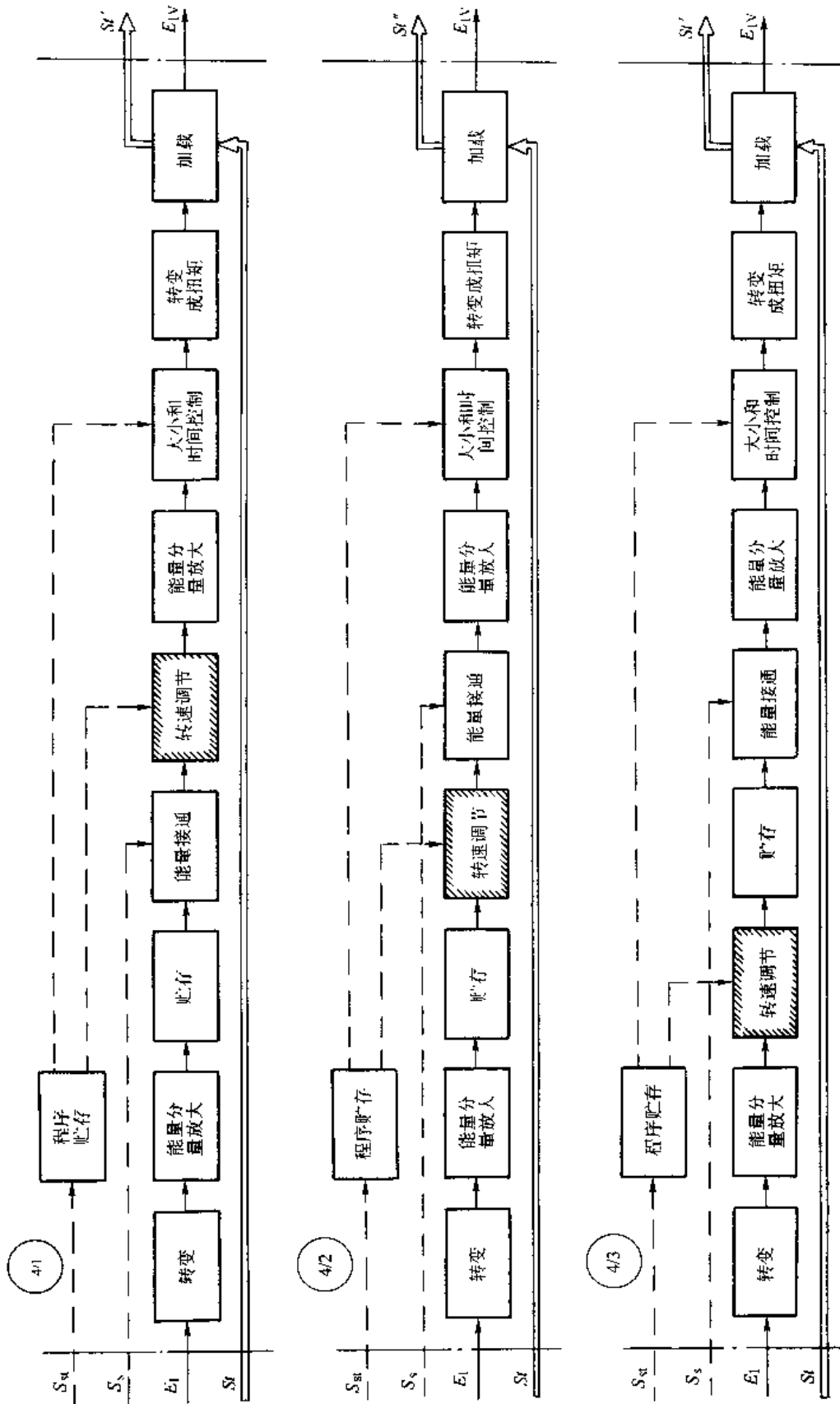


图 36.1-68 按照图 36.1-51 中功能结构 4 的功能结构变型

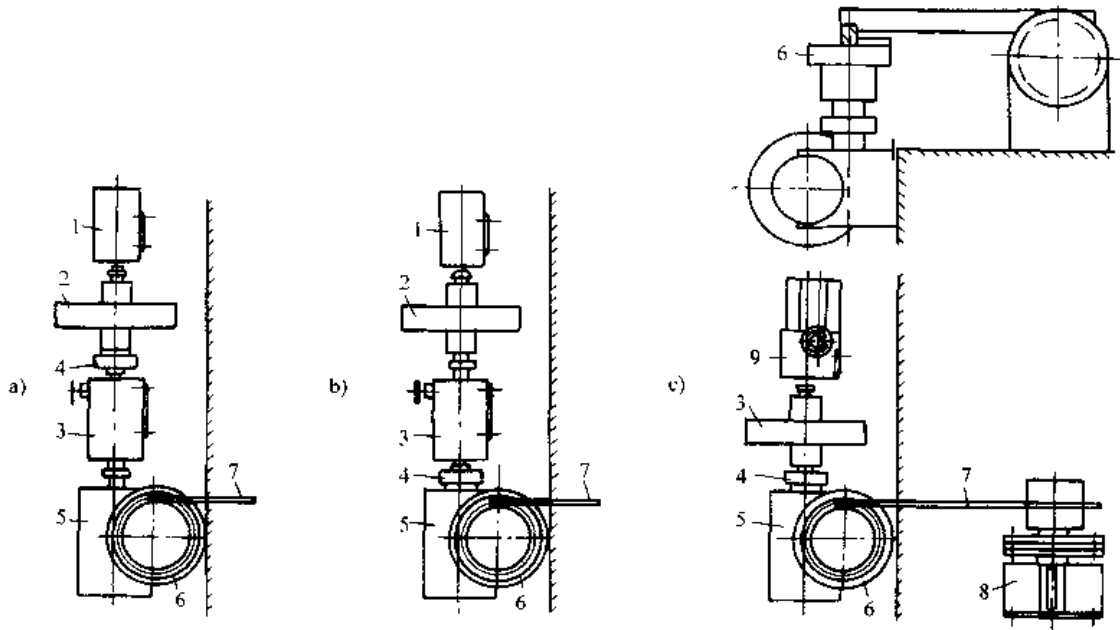


图 36.1-69 主功能载体的布置

a) 功能结构变型 4/1 b) 功能结构变型 4/2 c) 功能结构变型 4/3

1—马达 2—飞轮 3—调速减速器 4—可控式离合器 5—蜗轮减速器（锥齿轮减速器）  
6—凸轮 7—传力杠杆 8—试验联接 9—调速减速器马达

36.1-69c) 则采用了一个带调速减速器的马达。

**工作步骤 12:** 选择合适的技术设计方案

由于变型 4/3 采用了调速减速器马达（功能合成）而使得空间大大减小，因而在已设计出的结构设计变型中，选出变型 4/3 进一步进行具体化。

**工作步骤 13:** 其他主功能载体的粗略结构设计

其他主功能载体的粗略结构设计根据在工作步骤 11 中已经得知的下列要求进行：

— 凸轮驱动转速范围

$$n_K = 12.4 \sim 32.6 \text{ min}^{-1}$$

— 转速调节范围

$$R = 2.6$$

— 凸轮驱动力矩由

$$T_K = F_U \cdot D_K / 2 \quad \text{及} \quad F_U = F'_A \cdot \frac{T}{t_H} \quad \text{求得}$$

$$T_K = 2650 \text{ Nm}$$

— 凸轮驱动功率由

$$P_K = T_K \cdot \omega_K$$

得

$$P_K = 9 \text{ kW}$$

根据安全的理由，对于飞轮转速  $n_s$ （从而也是对于马达转速  $n_M$ ），选取最大值为

$$n_s = 1000 \text{ min}^{-1}$$

要求的传动比则为

$$i = 80.7 \sim 30.7$$

对于其他主功能载体，可对其特征值作如下估计：

— 离合器可传递的扭矩

$$T_{KU} = \frac{T_K}{i}$$

由凸轮的驱动力矩  $T_K = 2650 \text{ Nm}$  和凸轮与离合器之间的传动比  $i$ ，求得：

— 飞轮惯性矩

$$J_s = \frac{T_s \cdot \Delta t}{2 \cdot \pi \cdot n_s \cdot \Delta n}$$

由飞轮受到的扭矩  $T_s$  冲击时间  $\Delta t$ 、飞轮转速  $n_s$  和允许的转速差  $\Delta n = 5\%$  求得，

— 电动机功率  $P_M$  根据计算要求的加速力矩  $T_B$  算出。而

$$T_B = \frac{J_s \cdot 2 \cdot \pi \cdot n_M}{t_H} < T_{Bmax}$$

则由飞轮惯性矩  $J_s$ 、马达转速  $n_M$ 、起动时间  $t_H = 10 \text{ s}$  和由制造厂给出的马达最大加速力矩  $T_{Bmax}$  求得。

表 36.1-38 给出了主要特征量的计算值。

除了飞轮以外，这些主功能载体可以作为外购件获得，并且可作如下选取：

— 减速器

蜗轮减速器

功率  $P_G = 9.1 \text{ kW}$

输出扭矩  $T_G = 3800 \text{ Nm}$  ( $n_G = 32.3 \text{ min}^{-1}$  时)

传动比:  $i_G = 31$

— 电磁离合器

无环电磁离合器，型式 0-008-301，大小 17

表 36.1-38 功能结构变型 4/3 的与结构有关的主功能载体

功能	功能载体	特 征
能量转变 能量分量放大 转速调节	具有机械调速的电动机 - 变体 4/3	功率 $P_M = 1.1\text{kW}$ 转速 $n_M = 380 \sim 1000\text{min}^{-1}$ 转速调整范围 $R = 2.6$
能量贮存	飞轮	惯性矩 $J_S = 1.4\text{kg} \cdot \text{m}^2$ 转速 $n_S = 380 \sim 1000\text{min}^{-1}$
能量接通	电磁离合器	可传递力矩 $T_{KC} = 86\text{Nm}$
能量分量放大	传动箱	功率 $P_G = 9\text{kW}$ 输出扭矩 $T_G = 2650\text{Nm}$ (转速 $n_G = 32.6\text{min}^{-1}$ ) 传动比 $i_G = 30.7$

可传递力矩  $T_{K0} = 120\text{N} \cdot \text{m}$

- 调速马达

调速装置 R25-0000-075-4

功率  $P_M = 0.75\text{kW}$

转速  $n_M = 350 \sim 1750\text{min}^{-1}$

调速范围  $R \approx 5$

- 飞轮设计

转速  $n_S = 1010\text{min}^{-1}$

惯性矩  $J_S = 1.9\text{kgm}^2$

由于前面没有考虑摩擦等损失,  $J_S$  选得比以前规定的数值要稍微大些, 为了减少重量, 飞轮设计成空心的:

外径  $D_A = 480\text{mm}$  宽度  $B = 100\text{mm}$

内径  $D_i = 410\text{mm}$  质量  $m = 38\text{kg}$

按比例的粗略技术方案, 是在图 36.1-69c 中表示的主功能载体的基础上, 并且补充后面的主功能载体机架而作出的。

由于杠杆支承与试验联接一起的结构高度远远小于凸轮及整个驱动装置, 经与用户协商, 确定了试验台的空间边界条件, 见图 36.1-70。

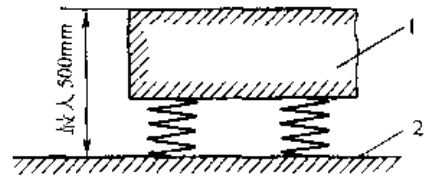


图 36.1-70 追加的空间边界条件

1—固定试验台的基板 2—地基

由于下述原因, 采用槽钢做机架的半成品:

- 以小的截面积得到大的面惯性矩,
- 无圆角,
- 有三个基面,
- 半成品价格便宜。

图 36.1-71 表示完成了的粗略技术方案。

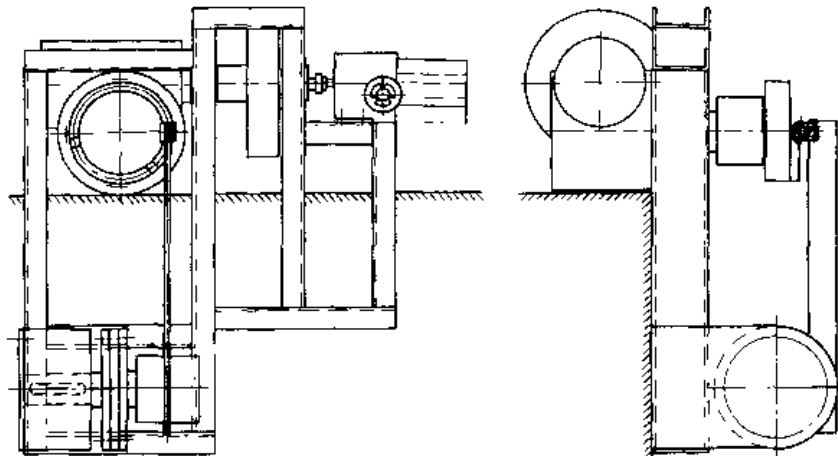


图 36.1-71 完成了的粗略技术方案

工作步骤 14: 寻求辅助功能解进行按比例的精  
确结构设计。

包括下述工作步骤:

- 寻找和选择辅助功能载体的解。
- 在辅助功能载体的基础上进行主功能载体的

精确结构设计。

- 辅助功能载体的精确结构设计。

这些工作步骤已不必要如同进行粗略技术设计时那样严格地分开了。由于具体化的程度高, 相互影响很大, 往往要求在更高的信息级别上将工作步骤重复。

辅助功能载体可以分为下列三类：

— 将主功能载体相互联接在一起的辅助功能载体。

— 将运动的主功能载体支承在机架上的辅助功能载体。

— 将主功能载体固定在机架上的辅助功能载体。

最后找到了下列解：

● 将主功能载体相互联接在一起的辅助功能载体为：

— 在杠杆与试验联接之间的法兰联接；采用成形联接的膜板联轴器，以避免附加的弯曲力矩，并且保证安装方便。

— 在蜗轮减速器与凸轮之间采用回转刚性联轴器。

蜗轮减速器与凸轮之间的联接可以采取两种方式，见图 36.1-72。

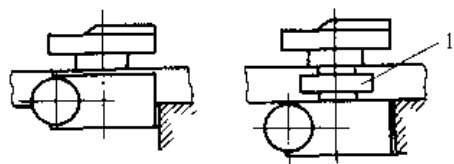


图 36.1-72 蜗轮减速器与凸轮之间的联接  
1—联轴器

带空心轴的蜗轮减速器——凸轮

蜗轮减速器——回转刚性联轴器——凸轮

采用回转刚性联轴器的理由为：

— 蜗轮减速器和凸轮可以分开装配

+ 便于装配的合理结构设计

— 由于轴的位置很高而不用断开机架

- 简单的结构设计

— 省去了为在蜗轮减速器与凸轮之间定心所需的较高的花费考虑工艺的合理结构设计，最后采用了 Arpex 回转刚性联轴器。

— 在飞轮与电动机之间采用回转弹性联轴器：Roflex 联轴器。

● 将运动的主功能载体支承在机架上的辅助功能载体：

— 飞轮支承

要求：飞轮加工简单（即不作价格昂贵的平衡），在吸收动态力方面的直接安全技术（保持安全原则），轴承在机架中悬挂布置。在这里不可能使用外购件（带滚动轴承的轴承座），因为通常铸造的轴承座主要是为座式工作而设计的。由于自生产飞轮产生的动态力大小不可确定，对轴承装置有必要进行专门的结构设计。

— 凸轮和杠杆的支承：市面上通常可以买到的

滚动轴承。

● 将主功能载体固定在机架上的辅助功能载体：

— 作为联接元件可采用简单的半成品（焊接薄板），将主功能载体用螺栓固定在上面。

— 专用的试验轮毅与杠杆（及机架）的联接要求：联接容易安装和拆开，轴向可移动，无间隙，公差不要太紧。

最后采用了环形弹簧夹紧装置 RfN7012。

工作步骤 16：考虑辅助功能载体的主功能载体精确结构设计

至此，主功能载体有必要与现已确定的辅助功能载体解相适应，分别来说，有：

调速马达：外购件

飞轮：图 36.1-73

— 联轴器：外购件

— 减速器：外购件

— 凸轮：图 36.1-74

— 杠杆：见总体技术设计草案（图 36.1-75）

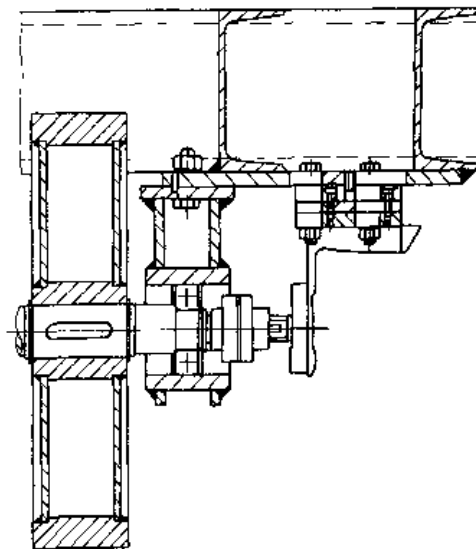


图 36.1-73 飞轮和飞轮轴支承  
(精确技术设计方案)

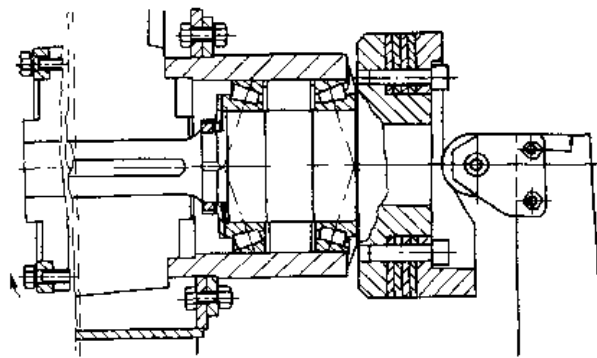


图 36.1-74 凸轮的支承 (精确技术设计方案)

——试验联接1见总体技术设计方案(图36.1-75)

——机架:由于马达结构型式改变,较之粗略技术设计方案有所变化

工作步骤10:辅助功能载体的精确结构设计和初步技术设计方案的完善

飞轮支承(举例):

——参数选择计算

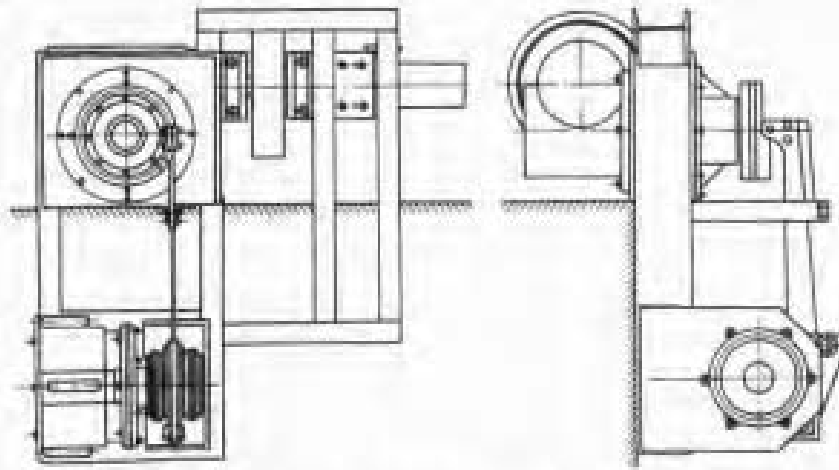


图 36.1-75 初步总体技术设计方案

轴承受力  $F_L$  估算如下:

$$F_L = F_g + F_d$$

其中重力为:  $F_g = m \cdot g = 400\text{N}$

动态力为:

$$F_d = m \cdot e_g \cdot 4 \cdot \pi^2 \cdot n^2$$

式中, 质量  $m = 40\text{kg}$

转速  $n = 1750\text{min}^{-1}$  (=最高马达转速)

飞轮偏心量  $e = 0.6\text{mm}$ , 其中:

尺寸和形状误差 0.3mm

轮-轴与轴承间隙 0.2mm

质量不均匀分布 0.1mm

得到轴承受力为

$$F_L = 1130\text{N}$$

这就是说,即使在出现离心力的情况下,轴承(额定动负荷 65000N)和所有处于力流中的零件的尺寸都是足够的。

——共振

轴承装置和机架的结构刚性很大,从而使得由飞轮产生的激振(最大 30Hz)处于其固有频率之外。

——加工

由于飞轮支承对机座没有紧的公差尺寸要求,因而得到工艺合理的结构设计。

——装配

通过以下几方面,使飞轮支承获得装配合理的结构设计。

——简单地由下面紧固。

——紧固螺栓具有拆卸方便的优点

在利用定位销对飞轮支承装置定位时,通过隔离

环使得电磁离合器容易调节(没有飞轮质量时本身也可能做到)。

——维护

通过使用免维护的轴承,获得考虑维护的合理结构设计。作为上述结构设计步骤的结果,图36.1-75给出了拟定的试验台初步总体技术设计方案。

工作步骤17:根据技术和经济准则进行评价

由于只作出了一种技术设计方案,因而评价的意义不是在于进行选择,而是按照具体的与要求有关的准则对现有设计进行判断,其目的在于认识和改善已有的薄弱环节,其进程可分成以下单个步骤。

——识别评价准则。

——判断在满足评价准则方面的特性。

——确定总价值。

——寻找薄弱环节。

对于找到的薄弱环节,在必要时提出改进建议。

作为评价的准则,基本上可以采用在方案设计评价目标系统中业已了解的重要准则。此时不是采用13个,而是11个准则,见图36.1-76,加权看来是不需要的了。

已设计出的试验台(图36.1-75)关键的或可计算的特性,可以以一个理想的解为基准,按照VDI2225利用从0~4分的粗略价值尺度进行评价。因为在评价上花费太多看来是不值得的。图36.1-76给出了评价的结果。

由于缺乏求得经济价值比的形式上可供使用的数据,因而在确定其总价值时只采用了技术价值比,  $W = 29/44 = 0.66$

由于其总价值较低,对其薄弱部位进行考察看来是值得的。

寻找薄弱部位通过找出分数低的特性而进行。对于只有1、2分的特性,在必要的情况下要立即作出改进建议。

- 操作失误可能性小
- 薄弱部位,马达转速。

转速必须要有可能调整得高于最大扭矩梯度所需要的数值。

由于温升的缘故,马达的启动只能在将其转速调整到转速范围的下限时进行。

补救措施:由生产厂提供转速计,标上启动和运行所允许的转速范围。

此外还可以在不允许的转速情况下自动关闸。

No.	评价准则		特性值		变型 4/3			变型 4/3 (修改)		
		加权		单位	特性 $P_{1i}$	价 $W_{1i}$	加权价 $W_{1i}P_{1i}$	特性 $P_{2i}$	价 $W_{2i}$	加权价 $W_{2i}P_{2i}$
1	可重复性		干扰影响	—	小	4				
2				—						
3				—						
4	超载可能性		超载程度	—	低	3				
5	工作安全性高		损坏危险	—	中等	2		见本文	4	
6	操作失误可能性小		操作失误可能性	—	高	1		见本文	3	
7	零件种类少		零件种类	—	低	3				
8	零件复杂度低		零件复杂度	—	低	3				
9	标准件和外购件多		标准件和外购件所占比例	—	高	4				
10	装配简单		装配简单度	—	高	3				
11	负荷变化方便		负荷变化	—	差	1		见本文	2	
12	试验联接更换快		试验联接的估计更换时间	—	中等	2		见本文	2	
13	测量系统可及性良好		测量系统的可及性	—	好	3				
		$\Sigma W_{1i} = 1.0$				$G_{4/3} = 29$ $W_{1i} = 0.66$			$G_{4/3} = 34$ $W_{2i} = 0.77$	

图 36-1-75 对图 35-1-75 所示轴承设计方案进行评价和寻找薄弱部位的评价表

- 加载过程变化方便

薄弱部位:由于杠杆对凸轮施加压力,使得不可能更换凸轮。

补救措施:采用杠杆提升装置。

- 工作安全性高

薄弱部位:回传的凸轮没有防护。

补救措施:采用防护栏。

- 更换试验联接迅速

薄弱部位:由于环形弹簧夹紧装置的可拆螺栓数目大,引起费用增高。

补救措施:不可能有补救办法,因为相应的形锁台联接要求很高的生产技术成本。

上面的补救措施已经记入了评价表(图 36-1-75)中。

图 36-1-77 表示完成了的冲击试验台。

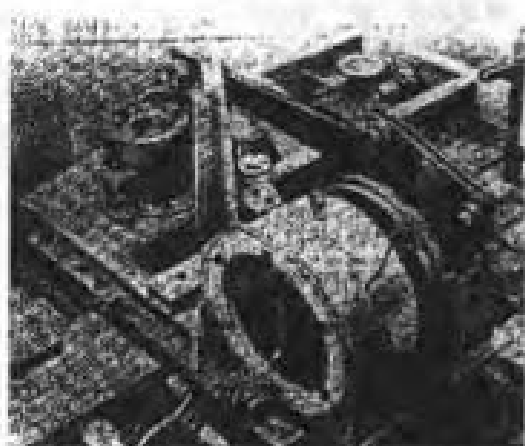


图 36-1-77 完成了的冲击试验台



## 第2章 产品设计中的“人-机-环境”大系统观

传统的设计观把思维集中在机械的功能实现和进行分析上,现代的设计观要求同时考虑人的因素和环境因素,以大系统的观点统一处理设计中的问题。

### 1 “人-机-环境”大系统的构成与运行

#### 1.1 “人-机-环境”大系统的构成

“人-机-环境”大系统的基本组成如图 36.2-1 所示。

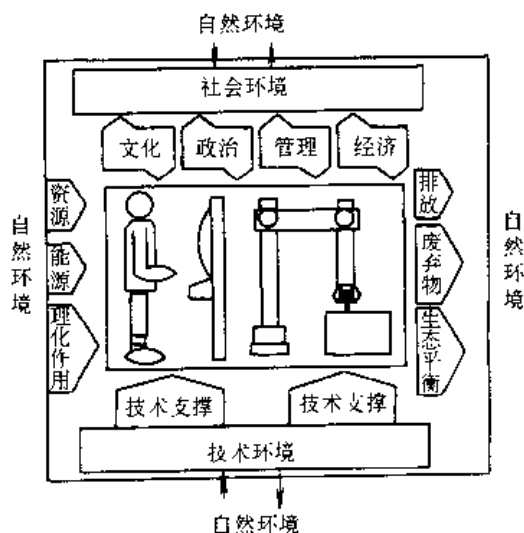


图 36.2-1 “人-机-环境”大系统

##### (1) 子系统——人

感觉器官、脑、效应器官是子系统“人”的3个组成部分,它形成人这个子系统的3个功能:

1) 刺激信号的接收 由眼、耳、鼻、舌、皮肤五个感官所产生的视觉、听觉、嗅觉、味觉、触觉来实现;

2) 信息的加工处理 所有接收的信息由传入神经传入大脑,经分析处理,形成决策,再由传出神经传给效应器官;

3) 效应器官的执行 效应器官由手与臂、脚与腿、口与舌、头与身等部分组成,它接收来自大脑的指令,并作出操作动作。现在人体生物电也成为效应器,由大脑支配的生物电信号已成功地用于假肢中。长期以来,手与臂、脚与腿是主要效应器官,随着用语言对话的计算机的出现,口与舌这一效应器也开始发挥重要作用。

##### (2) 子系统——机

由操纵、运转与显示三部分组成。

1) 操纵系统 是机接受人体效应器官操纵动作

的部分,它将人的操纵指令转化成机械运行的指令;

2) 运转系统 包括动力与传动系统、控制系统、执行系统等。控制系统对操纵系统接收的指令识别后,作出相应的动作,使机的运行状态作出改变;

3) 显示系统 机的运行状态改变将通过显示系统向人的感觉器官发出信号,使人得以判断机的新运转状态是否已满足要求。

简单机械常常不具备完善的显示控制系统。

##### (3) 子系统——环境

人与机都处于环境包围之中,并通过各自的界面与环境发生关系,在现代机械设计中,环境包括3个重要组成部分:

1) 自然环境 包括资源环境、生态环境和地理环境。从机向自然提取原材料起,经历日常运转,直到报废的全部寿命周期中,自然将不断地向机输入所需的物质与能量资源,并不断地接受机的排放与废弃物。人与机的共同行为将作用于包含人自身在内的生态环境,对生态平衡产生影响,而地理条件如气候、温度、湿度、风沙、日照、地形等等,将直接影响机的运行和人的劳动条件。

2) 社会环境 包括民族、文化背景、社会制度、政府政策、国际关系等方面。由于现代高科技产品通常都会给社会带来深刻的影响,因此上述社会因素也必然对产品的生产或使用发生促进或制约效果;由于现代产品大量参与国际大市场的竞争,因此市场环境成为产品开发的重要因素;由于产品的对象是人,因此人们的消费观念始终对产品的发展起导向作用。

3) 技术环境 包括设施环境和协作环境。现代化生产要求高度文明的劳动环境,它将由相应技术设施来实现。现代机械常常把群体的共性功能转交给公共的环境设施来承担,如大型客机的地面导航系统,船舶的卫星定位系统等。而像高速公路、加油站之类,则成为今天汽车运行的基础设施。现代机械的运作还需要大量的周边协作,如材料与燃料的供给、废弃物的回收等等。

### 1.2 “人-机-环境”大系统的运行

我们以汽车为例,来观察“人-机(汽车)-环境”大系统的运行情况及其相互作用。

#### (1) 人-汽车-环境”系统的基本运行

驾车人从车窗和后视镜看到前、后方道路,过往车

辆,行人等情况;还从路标、红绿灯以及车内收音机播发的交通信息台广播了解到更多的环境信息;并从仪表盘看到车速、贮油量等车况;然后通过脑的思维,作出决策,用手和脚操纵方向盘、刹车、油门、喇叭等操纵器,实现对车的驾驶。人还可以通过身体的振动感、耳朵听到的运转声,判断车的状态是否正常。

#### (2) 人与汽车

现代轿车的人-车界面宜人性不断提高。自动变速装置使车的驾驶变得十分容易;座椅的电动调节系统使人能以最适合自己的坐姿工作;装置在各部位的传感器使人通过仪表得以及时了解汽车的运行状态,及时对车进行维护保养;仪表盘的布置和仪表显示使驾车人一目了然;方向盘设计成可以伸缩及变换方位,使驾车人的手处于最佳位置,并且由于助力系统的存在,使驾驶变得轻快自如;严格按照人机工程学设计的乘坐空间,以及丰富的内饰件,充分保证了乘坐舒适性;高级音响设备不仅给人以享受,更重要的是为驾车人创造了消除疲劳的条件;人车关系在现代轿车中越来越趋融合。

#### (3) 人、车与自然环境

汽车是全天候交通工具,运行环境变化极大。在气温 $40\sim 50\text{C}$ 的热带,车上的橡胶件将发粘,在 $-40\text{C}$ 的寒冷地带,橡胶件又会变脆,因此,采用既耐高温,又耐低温的车用橡胶材料;为防止油箱中润滑油冻结,在润滑油内应加防冻添加剂;在沙漠地带运行时,由于进气中灰沙含量大,会使汽车发动机磨损加剧,必须增加保护装置;为解决雨天行车时雨水在车窗上遮挡视线问题,车窗上装有刮水器;为解决在雪地行车时车轮打滚问题,车轮上加装防滑链;而为了保证人在各种自然环境下有一个舒适的乘坐环境,现代轿车都装有空调器。

人和车反过来对自然环境也产生很大影响。长期以来,排气污染一直由大自然默默地承受,但到了今天,污染已成为人类巨大的灾害,迫使人们不得不通过技术和立法,改善自身行为。在发达国家轿车中普遍采用燃油电子控制喷射装置、无铅汽油、尾气净化装置三大措施,对轿车排放的尾气进行控制。正在研制的电动汽车和其他燃料汽车将进一步解决人类大量用车对自然的侵害。

汽车的广泛使用与自然资源的承载力之间的矛盾也日益尖锐,据目前勘测,世界石油储备量只够维持人类40年的消耗。人类不合理地追求高消费,要求汽车不断变换车型,大量材料资源被消耗用来制造新车,同时旧车又成为工业垃圾对环境带来沉重的负担。现实迫使人们冷静下来,重新考虑汽车的消费方式,并产生可回收、可再生利用的设计技术。

#### (4) 人、车与社会环境

汽车的出现改变了人类生活,在一些发达国家,没有汽车感到寸步难行。美国被称为是“架在汽车轮子上的国家”。汽车工业在很多国家已成为带动国民经济发展的支柱产业,汽车市场的竞争成为国际贸易关系中的热点,由此产生出很多国际间的汽车贸易法规。然而,汽车又成为社会的一大公害,每年有数百万人死于车祸。在欧洲甚至出现了反汽车运动,要求至少在市内用自行车代替汽车。当然,人们更多的是从积极方面解决问题,例如,鉴于酒后开车是事故的重要起因,现已研制出对司机的呼气中酒精含量在车监测装置,若呼气中酒精含量超出规定值时,该司机将无法启动汽车。为保护驾车人在撞车事故中免受致命伤害,发明了安全气囊。这种气囊在汽车一旦相撞时,会立即充气,在人与方向盘之间起缓冲作用,避免坚硬的方向盘压人驾车人的腹部与胸部。另外,一些防止汽车压死人的安全措施也在研究之中。

社会制度对汽车的发展起着一定的作用。我国以往一段时期曾禁止私人购车,轿车工业几乎是象征性的,现在汽车工业被政府定为支柱产业,轿车成为发展重点,私家车拥有量正不断上升,轿车的自主开发从此将会取得长足发展。汽车市场销量在我国也受政府政策影响极大,通过价格对市场进行了严格的控制,未来家庭用车的发展也基本上取决于政策。

文化背景常给汽车设计带来很大的影响。在德国、桑塔纳轿车是针对普通家庭设计的,前排为主人翁乘坐,后排安置小孩和狗。这项设计全盘照抄到中国就很不适应,为此作了加长改型设计。从法国雪铁龙引进的富康车型,采用车厢与行李舱融为一体的车身设计,具有极好的空气动力性能,却因外形“没有尾巴”,在一些地区人们心理中产生“断后”的不祥心理,一度影响销售。今天,汽车造型已成为社会时尚,新车型推出的速度已缩短到平均两年,轿车档次的高低还成为车主社会地位的象征。

#### (5) 人、车与技术环境

现代汽车的技术水平与道路有着密切的关系。时速达 $150\text{km/h}$ 以上的汽车是完全针对高速公路而设计的;为减轻汽车行驶时噪声的影响,在城市高架道路两侧装置了降噪墙;各种交通指示牌向司机显示环境信息,引导驾驶,其良好的夜间泛光显示性能给司机夜间行车带来很大方便。目前,世界各国在实施左右驾驶制度上因习惯不同而不统一,如中国采用左驾驶制,方向盘设于前排左方,而日本采用右驾驶制,虽然它给汽车生产带来极大的麻烦,但各国大量的交通设施都随此而定,较难更改。

随着汽车技术的进步,科技人员将能把现在汽车

的相当部分功能转移给环境设施来承担。人若有较长时间在汽车里度过,就要发展无人驾驶汽车,它的前提是建立卫星导航系统和道路的导航设施。到那时,人们只要上车后设定终点目标,汽车将自动行驶到目的地,人们在车上可继续工作或休息。为解决环境污染,各国都在发展天然气汽车和电动汽车,但与此同时必须建立供气或充电网点。未来汽车的运行将越来越多地依靠环境设施的配合。

就在汽车运行的背后,很多与汽车相关的科学和技术正在协同运行着。为消化汽车运行中的废弃物,大量相关交叉科学技术被应用,为将废轮胎作为一种资源加以再利用,就产生了一系列相关的科学和技术,如再生橡胶技术、用聚胺酯修复技术、用细菌分解转化废橡胶为其他材料技术、用废旧轮胎提炼石油技术以及将废轮胎转化为高能燃料技术等。

## 2 功能分配

### 2.1 子系统能力的分析

#### (1) 人

人通过感觉器官十分敏感地接受综合信息并将这些信息通过大脑进行分析和决策。

但是人的能力也受到多种因素的限制:

- 1) 信息通道小,存储信息数量有限;
- 2) 感觉器官信息接受能力有限;
- 3) 受生理和心理因素影响,功能发挥可靠性差;

1) 对环境敏感。

(2) 机

1) 机能远远超出人的生理和心理极限,发出工作所需要的力和功,达到所需要的速度,并长时间运转;

2) 工作稳定,抗干扰能力强,对环境的承受力强;

3) 可以在更大的范围内感知信号,如感知紫外线与红外线、次声与超声波、极微小的变化等,感知速度也快。

4) 内存信息量大,通过计算机系统可记忆大量数据资料,运算速度快。只要对某项处理过程中人的思维模式掌握清楚,由机代替人进行处理,能力明显加强。

人与机能力的详细比较见表 36.2-1。

表 36.2-1 人和机器要素的特性比较表

比较项目	人的特性	机器的特性
检测能力	<ol style="list-style-type: none"> <li>1. 灵敏度随时间变化,易受环境和感情的影响</li> <li>2. 有主观倾向性</li> <li>3. 虽然味觉和嗅觉无法定量表示,但比机器灵敏度高</li> <li>4. 难以监测偶然发生的事件</li> <li>5. 只能单通道</li> <li>6. 能在高噪声环境下检出需要的信号</li> </ol>	<ol style="list-style-type: none"> <li>1. 固定性良好,稳定</li> <li>2. 灵敏度、准确度、检测范围良好</li> <li>3. 重复性好</li> <li>4. 线性良好者多</li> <li>5. 监控能力很强</li> <li>6. 专用、缺乏通用性</li> <li>7. 能够进行多通道的复杂动作</li> <li>8. 无错觉</li> <li>9. 在高噪声环境下很难正确无误地接受信号</li> </ol>
操作能力	<ol style="list-style-type: none"> <li>1. 由于熟练而能发挥机器所不能及的巧劲</li> <li>2. 训练一般需要较长时间,有时短时也可训练好,若训练不妥会成为错误操作的根源</li> <li>3. 能进行高精度操作和复杂的艺术性工作</li> <li>4. 10s 内能输出 1.5kW 功率,以 0.15kW 的输出能连续工作 24h,并能精细调整</li> <li>5. 通用</li> </ol>	<ol style="list-style-type: none"> <li>1. 重复性、准确度、线性度良好</li> <li>2. 能输出极大和极小的功率</li> <li>3. 不能像人那样进行精细的调整</li> <li>4. 反应快、效率高</li> <li>5. 专用</li> </ol>
判断能力	<ol style="list-style-type: none"> <li>1. 有智能、有特殊判断能力</li> <li>2. 人与人之间信息交流效率低</li> <li>3. 需要训练,否则易出错误</li> </ol>	<ol style="list-style-type: none"> <li>1. 高速</li> <li>2. 定量记忆良好</li> <li>3. 能同时高效协调地运转两个以上装置</li> </ol>
感觉能力	<ol style="list-style-type: none"> <li>1. 可识别物体形状、大小、位置和颜色</li> <li>2. 对不同音色及某些化学物质有一定的分辨能力</li> <li>3. 能感觉到最低强度的光和声</li> <li>4. 对各种刺激的感觉比机器灵敏</li> <li>5. 能通过直觉从许多目标中找出真正的目标</li> </ol>	<ol style="list-style-type: none"> <li>1. 识别图形能力差</li> <li>2. 有接受人所不能感觉的刺激,如接受红外线,超声波、辐射、微波、电磁波、磁场等信号的能力,并做出记忆和反应</li> <li>3. 能对给出的信号迅速做出反应</li> <li>4. 只能发现特定的目标</li> </ol>
信息处理能力	<ol style="list-style-type: none"> <li>1. 可接受语言和表情等信息,但容量少</li> <li>2. 接受信息速度慢,每秒约 20 个</li> <li>3. 每次只能处理一个信息</li> <li>4. 传递信息速度一般为 6bit/s</li> <li>5. 工作记忆容量小 (5bit),存取时间 0.025s,长期记忆容量很大 (<math>10^6</math>bit),存取时间 0.2s</li> <li>6. 反应时间最小值为 <math>10^{-3}</math>s</li> </ol>	<ol style="list-style-type: none"> <li>1. 微机主机容量很大外存容量无限,存取速度快</li> <li>2. 能长期储存信息,短时期内能储存和回忆大量信息</li> <li>3. 对信息传递能力、记忆速度、保持能力比人强很多</li> <li>4. 能同时处理多种信息</li> <li>5. 能以各种形式表示输出的信息</li> <li>6. 反应时间为微秒级</li> </ol>

(续)

比较项目	人的特性	机器的特性
控制能力	<ol style="list-style-type: none"> <li>1. 可进行各种控制,在自由度、调节和联系能力等方面均优于机器</li> <li>2. 本身具有各种运动的动力</li> </ol>	<ol style="list-style-type: none"> <li>1. 操纵力、速度、精确度、操作频率等方面均优于人</li> <li>2. 需外加动力才能发挥作用</li> </ol>
创造能力	<ol style="list-style-type: none"> <li>1. 具有创造性,能对各种问题具有全新的见解,具有发现特殊原理或关键措施的能力</li> <li>2. 有思考、分析、归纳、推理的能力</li> <li>3. 有根据经验改变操作的能力</li> <li>4. 对意外事件有作出反应的能力</li> <li>5. 有随机应变能力</li> <li>6. 对事物的发展能做出相应的预测</li> </ol>	<ol style="list-style-type: none"> <li>1. 没有创造力,但可以在程序功能的范围内进行一定的创造性工作</li> <li>2. 只能处理已知的事件</li> <li>3. 无随机应变能力</li> <li>4. 预测能力有很大局限性</li> </ol>
耐久性	<ol style="list-style-type: none"> <li>1. 不能长时间连续工作,易产生疲劳,有短期内超负荷工作的能力</li> <li>2. 记忆力、熟练程度受年龄、性别、健康状况等因素的影响</li> <li>3. 有自我恢复能力</li> <li>4. 不善于从事单调工作,不能自发保持紧张</li> </ol>	<ol style="list-style-type: none"> <li>1. 耐久性好,保养良好可长期使用</li> <li>2. 能长期连续、快速、精确地重复工作,而不像人那样会产生疲劳</li> </ol>
可靠性	<ol style="list-style-type: none"> <li>1. 错觉多</li> <li>2. 决定于熟练程度</li> <li>3. 发生紧急情况时情绪会影响可靠性</li> <li>4. 技术水平、生理状态、心理状态、社会因素对可靠性有影响</li> <li>5. 有善于排除故障、处理意外情况的能力</li> </ol>	<ol style="list-style-type: none"> <li>1. 决定于设计质量、材料、加工、装配、维修保养质量</li> <li>2. 本身无维修能力</li> <li>3. 一般不能处理意外的紧急情况</li> <li>4. 一般比人可靠</li> </ol>
适应性	<ol style="list-style-type: none"> <li>1. 可依次完成多种功能作业,但不能同时完成多种作业</li> <li>2. 不能在恶劣环境下工作</li> <li>3. 不能进行高阶运算</li> <li>4. 具有很强的学习能力,能阅读和接受口头指令,灵活性很强</li> </ol>	<ol style="list-style-type: none"> <li>1. 可同时完成多种作业</li> <li>2. 进行单调而重复工作不会降低效率</li> <li>3. 能进行高阶运算</li> <li>4. 能耐恶劣的环境,能在放射线、尘埃、有毒气体、高噪声、黑暗、强风大雨等条件下工作</li> <li>5. 能进行笨重的工作</li> </ol>
其他	<ol style="list-style-type: none"> <li>1. 成材的时间长,约20年</li> <li>2. 人的智能可不断开发和提高,但也能退化</li> <li>3. 受社会环境、家庭生活等因素的影响</li> </ol>	<ol style="list-style-type: none"> <li>1. 可大量生产同一规格产品</li> <li>2. 功能一般可长期保持不变</li> <li>3. 能24h连续工作</li> <li>4. 不受社会条件的影响</li> </ol>

### (3) 环境

它通常从如下几方面承担自己的功能:

- 1) 在系统运转过程中不断提供物资与能源。
- 2) 接受系统的排放,并予以消化。
- 3) 创造人机良好工作的周边环境。工作场所的文明度对人的情绪和劳动质量影响很大,合适的温度和照明是人最基本的工作条件。为保证机的良好工作状态,通常也必须创造相应的工作环境。
- 4) 用环境设施来承担机械群体的某些共性功能。如机场内的导航系统就是把飞机自动驾驶装置中一部分功能由飞机转向地面环境设施的系统,它为所有在航线上飞行的飞机所共享,这样做明显比在一架飞机内孤立地设置全套自动驾驶系统经济合理,而且可靠。
- 5) 从社会角度解决系统运行问题。如制定行业标准、规范,制定国际标准。

## 2.2 功能分配原则

1) 比较分配原则 对人、机、环境的特性作比较,据此进行客观、逻辑的功能分析。将大系统的各功能分

配给适合的子系统承担。

2) 剩余分配原则 即把尽可能多的功能分配给机,尤其是计算机。剩余的功能分配给人,因为人将承担更有意义的工作。

3) 经济分配原则 以经济效益为根本依据,一项功能分配给人还是机或环境,完全视经济分析来评定。

4) 宜人分配原则 把人的利益放在首位;要把人从繁重的体力劳动中解放出来;要顾及人的承受限度;要使人仅承担最关键的工作;要考虑人比较容易上手,以降低培训消耗;要让人的物质与精神文明需要从中得到最大的满足。

5) 弹性分配原则 即由人在系统运行中自己选择和参与系统运行的程度。如民航客机设有手控和自动飞行两套操纵系统,一般情况下自动飞行,关键场合由人控制。

总的趋势是:

- 1) 使人不断从繁重的体力劳动中解脱出来,在大系统中更多地发挥智力强项的功能;
- 2) 使人的一部分有明显规律性的思维转由机去

执行,即由计算机或自控系统执;

3) 系统的宜人要求日益受到重视,人不仅承担系统分配的功能,而且是物质和精神文明的享受者;

4) 机对环境设施的要求也越来越高,很多机械需要公共设施的支持,与社会环境的关系也日益紧密;

5) 人们已从机的全寿命周期考虑“人-机-环境”大系统的运转,以及贯穿各阶段的功能分配;

6) 自然环境供给资源和接受排放的功能应符合可持续发展战略。

### 3 人的生理和心理因素

#### 3.1 人的生理因素

在机械领域的设计中,包含如下相关因素:

1) 人体的几何因素 包括身高、体形、肢体尺寸及比例等,他们随性别和人种而异,是人与机匹配设计时的重要依据。

2) 人体的力学因素 人既能作微细的调节动作,也能发出很大的力,但人的出力有一定的限度,并随年龄、性别、作业方位和环境的不同而不同。

3) 人的持久性 人如果长久地工作,特别是以固定的姿势、作固定的动作时,肢体、感官和大脑都将会产生生理和心理性疲劳,出现如体力下降、睡意上升、注意失控、思维迟钝、工作意志衰退等现象。须经过一段时间休息,才能恢复正常。自动化的目的之一正是克服人的这一不足。

图 36.2-2 是企业中工人在二个工作日的 8h 之内劳动能力变化曲线。

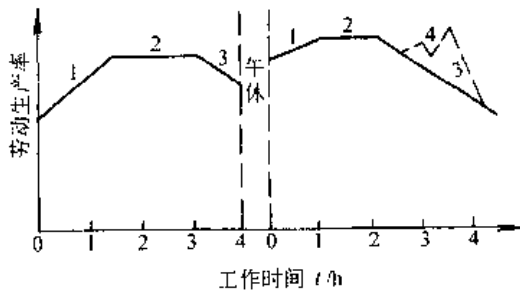


图 36.2-2 工作日内工作能力的变化

4) 人的生理节奏 人是在自然界中生存和发展的,昼夜变换对人的影响极大,形成日生理节奏。人的体力、情绪和智力存在着 23d、28d、33d 正弦周期变化规律。此外,女性还存在月生理周期。

图 36.2-3 表示一天中人的 CFF 值变化曲线,以 0 时的测值为基准,从中可见 6 时 CFF 值最低,中午 12 时最高。

5) 人的环境承受力 人适应工作环境的能力是十分有限的。人对温度、光亮、声响、振动、冲击、气

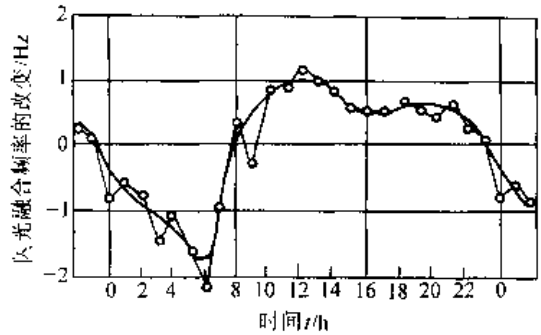


图 36.2-3 闪光融合频度的昼夜变化

压与水压、有害气体与辐射等方面的承受力远比“机”小。

6) 劳动强度 人的短期迸发力与长期负荷能力的差异是很大的。劳动强度反映人在一个完整的劳动过程中平均体力或脑力消耗,通常用心率、搏出量、需氧量等指标来衡量。

心率指心脏每分钟搏动次数,搏出量为每次搏动从左心室射入主动脉的血量。正常人在安静状态时的心率为 60~100 次/min,搏出量为 60~80ml/次,均随个体和年龄而异。因劳动强度不同,心率发生很大的变化,一般平衡状态心率 100~125 次/min 为中等劳动强度,125~150 次/min 为强劳动,175 次/min 以上为极强劳动,搏出量在劳动中将增加 40~80ml/次。体力强、经常参加锻炼运动的人,心率与搏出量随劳动强度变化较小。

人体为维持生理活动,体内所需氧气的量称为“需氧量”,成人安静时约为 200~300ml/min。劳动强度与需氧量关系见表 36.2-2。

表 36.2-2 按能耗和需氧量分级的劳动强度指标

劳动强度级	轻	中等	强	极强	过强
能耗下限/ kJ·min <sup>-1</sup>	10.5	20.9	31.4	41.9	52.3
需氧量下限 L·min <sup>-1</sup>	0.5	1.0	1.5	2.0	2.5

由毛细血管实际摄取氧气的量叫做“摄氧量”。人的摄氧能力是有限度的,当需氧量大于摄氧量时,会造成体内氧亏损,称为“氧债”。一旦出现氧债,人体肌肉组织内产生乳酸等成分,乳酸需通过体息才能消除。人的氧债承担能力是有限度的,一般人为 10L,像运动员等经常锻炼者可达 15~20L。氧债量的加大将使血液中乳酸量大量积累,形成抑制作用,产生疲劳不支的感觉,以强迫人体息,起到自然的生理保护作用。

#### 3.2 人的心理因素

按照心理学理论,人的心理因素可概括为动态的心理过程和静态的心理特征两大方面。

### (1) 心理过程

心理过程包括认识过程和意向过程。

1) 认识过程 由客观到主观的过程,包括人的感觉,知觉和思维过程。

**感觉** 是人通过感觉的直接体验,在心理学上分为三大类,即:

① **外部感觉** 通过人体外部各种感官接收到的外部刺激和信息,如光、温度等;

② **内部感觉** 通过人体内部的神经末梢接收到的肌体内部刺激和信息,如饥饿、口渴、腹痛等;

③ **运动感觉** 来自肌肉的运动感。

**知觉** 不是直接的感受,而是客观事物在人脑中的反映,它可划分为:

① **时间知觉** 对客观事物延续性、顺序性的反映;

② **空间知觉** 对客观事物形状、大小、远近、方位等空间特性的反映;

③ **运动知觉** 对位置、移动、速度快慢的反映。

**思维** 是人类的最高级的大脑活动,它包含分析、综合、比较、抽象、概括、具体化等基本过程。它的基础是记忆,人能够把主体经验过的事物、感受、情绪、动作、语词概念等大量的信息加以识记和保持,在必要时加以再识或回忆,构成“记忆”的完整心理过程。

2) 意向过程 由主观到客观的过程,包括人对待与改造客观的情感、意志和对事物的注意等内容。

**情感** 客观现实和人的需要之间关系的态度体现,是一种复杂的心理功能,并又和很多其他的心理功能密切联系。

情感的对外表现是情绪,在“人-机-环境”大系统中,情绪是十分重要而又极不稳定的因素。情绪可分为三类:

① **心境** 是持久的、弥散的情绪状态。处于某一心境的人,会以同样的情绪对待一切事物。设计者应考虑为人营造一个保持良好心境的环境。

如:客机座舱,轿车内饰设计,现代化车间设计等

② **激情** 是一种迅猛暴发、激动而持续时间短暂的情绪状态,有些“机”必须具有激情的人配合使用,这时的设计重点就不是保持良好的心境,而是激发人的激情。

如:赛车的造型,要使车手一看就产生一种高速、冲刺的欲望。

③ **应激** 是在出乎意料的紧急状态下所引起的情绪状态。这时人的思维、行动和整个机体都高度激化,能做出平日较难做的事。

如:飞机紧急逃生气囊式滑梯就是按应激状态设

计。

**意志** 是人们自觉地确定目的,并根据目的来支配,自觉地调节自己的行动,去克服困难以实现预定目的的心理过程。追求目的和努力克服困难是意志的两大特征。

如:现代产品设计要考虑使用者的学习意志,使用简单方便的产品易于推广。

**注意** 是人的心理状态对一定对象和活动的指向和集中。在设计中应考虑4个方面:

① **注意的范围** 是指在同一时间内注意所把握对象的数量和空间,被注意的对象越集中,排列越有秩序,就越有利于人们把它们作为相互联系的整体来知觉,注意的范围将因此而扩充。

② **注意的分配** 是指在同一时间内把注意力分配到两种以上不同对象或活动上。

例如,汽车司机在开车时,既要注意前方路况、车辆和行人,又要注意手中转向盘和脚下的油门、刹车。人们会自动分配自己的注意力,把注意集中到自己感到生疏的事物上。对注意作过多和分散的设计是引发事故的主要原因。

③ **注意的稳定性** 是指在一定的时间内把注意投放在特定事物上的持久性。对于贫乏、单调、不变的对象,人的注意力将很难保持持久。

④ **注意的转移速度** 人的注意力转移速度是有限度的,频繁、过快的转移会使操作者感到疲劳。

### (2) 心理特征

心理特征是人的气质、能力和性格的综合,既包含天赋,更包含后天学习和社会实践的结果。

1) **气质** 在进行心理活动时或在行为方式上表现出来的强度、速度、稳定性和灵活性等动态性的心理特征,是人的特性类别,先天成分较多,但无优劣之分,如人的内向与外向,好静与好动,敏捷与迟钝,优柔与果断,容易激动与善于控制等。各种气质的人都有适合自己的工作:仪表装配线上要选好静的女工,司机应是敏捷型的人,规范性很强的工作宜挑选安守本分的工作人员,赛车手必须是富有激情的人。

2) **能力** 成功地完成某种活动所必需的个性心理特征。包含先天和后天、生理和心理几方面的成分。如设计人员为女性设计摩托车,就应该优先考虑小轮、前方上车、轻巧等特点,以适应女性的生理与心理方面的承受能力。现代机械产品对人的能力要求在向两极发展,一些而向广大消费者的产品,设计者把越来越多的功能分配给“机”去承担,对人的能力要求尽量降低;一些高技术生产型产品则对人的能力要求日益提高,如加工中心的操作。人的能力可通过培训来提高,许多产品在开发的同时,销售部门应同时考虑用户培训问

题。很多生产高科技产品的人公司,它的培训网点往往是全球性的。

3) 性格 表现人的态度和行为方面比较稳定的个性心理特征,在个性中起核心作用,有先天、后天两方面成分,有好坏之分。勤劳、诚实是企业职工的好性格;懒惰、虚伪的职工是企业的灾害。

### 3.3 人的行为

行为是心理过程支配下的动作过程。外界的刺激作用于人,通过感知和思维,形成判断和决策,通过效应器官执行,构成动作。面向某事物而产生的相关动作的组合,则构成行为。

行为科学的很多研究结果已成为现代机械设计的重要知识,特别是消费行为、操作行为和事故行为。

#### (1) 消费行为

行为意识来自动机,动机来自需要,人的需要由低级到高级按图 36.2-4 所示不同层次排列。处于不同消费层次的人,具有不同消费心理和消费行为。



图 36.2-4 需求的不同层次

经济始终是影响消费行为的重要因素,经济实力不足时会抑制消费,而经济富裕时则易滋生奢侈浪费和不科学的高消费。

#### (2) 操作行为

人通过感官从被操作对象处获得信息,经过思维作出动作决策,由手、脚或人体某些部位执行,形成操作行为,其中手的动作最为重要。

1) 操作过程的动作分析 分析的目的在于建立科学的操作过程,或制定标准的操作时间,它包括 3 个方面:

- ① 确定必需而有效的动作要素;
- ② 排除无效的动作及诱发原因;
- ③ 减少影响效率的辅助动作。

2) 操作过程设计的动作经济原则 其要点在于设计一套操作环境舒适、操作劳动强度低、时间短、效率高的操作过程和方法。它可以归纳为如下 4 个基本原则:

- ① 争取两手同时动作,或把动作分配给占优势的肢体。如一般人的右手为优势手,食指为优势指,其

动作最准确、灵活;又如腿的出力最大;

- ② 力求减少动作单元;

③ 力求缩短动作距离,减少动作加、减速引起的能量消耗;

- ④ 建立正确的作业姿势,提高操作的舒适性。

3) 规范的操作行为设计 建立科学的、严格的操作规程并对操作者进行必要的岗位培训,建立违章操作惩罚制度,是规范操作行为的重要措施。

#### (3) 事故行为

在“人-机-环境”大系统中,机的设计不当、环境条件的变化都会引发事故,人是其中最不稳定的因素。据统计,80%的事故是由于人的原因造成,在特定条件下人作出了引发事故的行为。这些原因可分为局部和全程两种。

##### 1) 局部事故行为

① 感知觉障碍 如机和环境传给人的信息刺激太弱,或同时发生的信息太多、太乱,使人不能准确接受,出现漏读、错读仪表;或忽略了某些环境现象观察,或彩色显示的系统误用了色盲患者,以及各种因素造成的注意力分散等。

② 思维决策障碍 如决策所需的知识不够,缺少某些该记忆的资料,思维速度跟不上系统运行速度,疾病或酗酒引起的思维障碍,性格的优柔寡断等

③ 行动障碍 如操作人员选用不当,体力达不到操作力的需要,或身材够不到操作位置;多项操作位置彼此距离太远,不能保证操作人员及时到位等。特别是一些必须时刻处于人操纵状态的系统,应精选体力充沛的人员上岗。

##### 2) 全程事故行为

① 思想情绪 异常兴奋的情绪将导致行动轻浮、散漫;忧郁的情绪将导致感知与思维的迟钝,这些都会引起操作失误。

② 觉醒水平 受情绪和生理心理节律的影响,正常的生理性疲劳,对从事工作的厌倦,这些都会影响觉醒水平。高水平的设计应使工作安排与节律匹配,特别是在与自然节律相隔离的工作场所。

③ 能力冗余 经过反复训练的操作人员能实现快速、熟练、下意识的操作过程,使操作人员能力出现冗余,导致注意力外移从而引发事故。设计者应考虑熟练工人有可能的意外行为,必要时采取保护措施,如冲床中设置误操作保护装置等。

④ 过分紧张 因生产节奏太快,或动作不熟练而产生恐惧,引起的过分紧张,这种状态将对感知、思维和动作过程发生不利影响。

⑤ 工作单调 重复单调的操作会使人失去劳动的创造心理,影响劳动情绪,提前产生心理疲劳或注意

力分散。在人力所及的限度内,扩大充实操作内容,充分调动人的积极性。定期变换工作内容或岗位也能带来同样积极的效果。

⑥ 周边环境 主要指操作环境,如温、湿度、照明、操作空间、振动与噪声、空气、文明程度等。舒适的环境对情绪、觉醒程度、工作耐久性有很大影响。

⑦ 人体疲劳 包括生理疲劳和心理疲劳。它使操作者从感知觉开始的整个操作过程的动作变得迟钝或最终失控。

⑧ 缺乏训练 详细了解可能发生的事故以及防止的方法,预先对事故排除过程进行模拟演习,这对减少事故率、减轻事故的后果有很大作用。

#### 4 环境因素与可持续发展观

##### (1) 可持续发展战略

可持续发展观强调的是环境与经济的协调发展,追求的是人与自然的和谐。它包含如下一些内涵。

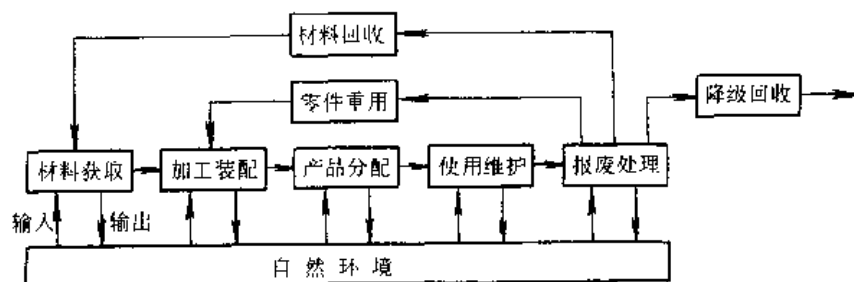


图 36.2-5 产品的生命周期与环境的关系

1) 原材料获取阶段 在开采制备阶段,所选择的应该是能耗低、成本低、污染少、无毒性,或经过处理,毒性可以消除的材料,对可再生资源应考虑再生周期和成本,对不可再生资源则应严格控制用量,特别是稀有资源应保护性使用。这一切,当相应的资源法规健全后都将会反映到材料成本中;在加工过程中,所选材料应易加工,低能耗、废料不污染环境;在使用中不应产生有害物质,在产品寿命终止后,应易降解和回收。

2) 加工装配阶段 对污染采取主动防治的技术路线,争取减少或消灭在生产过程的上游,特别是在产品加工制造阶段产生的污染。这样不仅正本清源,从根本上保护环境,而且给企业带来经济效益。

3) 产品分配阶段 这一阶段主要涉及产品包装和运输。要求产品具有良好的可运输性,节约空间,减轻自重,以降低运输过程中的能源消耗和对环境的污染。

4) 产品使用阶段 产品在使用阶段的能耗、资源消耗、排放物以及易维护性,是可持续发展论关心的热点。

5) 产品报废阶段 在可持续发展观念中,报废产

1) 可持续发展的前提是发展。

2) 可持续发展的核心之一是协同。即社会进步的目标、经济增长的目标、环境保护的目标三者之间的协同,也即人类社会与自然环境的协同发展,要使人类的发展有一个可持续的资源保证和协调的环境支持

3) 可持续发展观的核心之二是强调代际之间的机会均等。指出当代人享有的正当环境权利,即享有在发展中合理利用资源和拥有清洁、安全、舒适的环境权利,后代人也同样享有这些权利。

##### (2) 产品生命周期中的可持续发展观

产品的生命周期包括原材料获取、零件加工和产品分配、产品使用和维护、产品报废等阶段。各阶段都需要有物料和能源输入,也都有废弃物向周围环境输出,如图 36.2-5 所示。

可持续发展的思想应贯穿到产品生命周期的每一个环节。

品始终是资源,应进入再循环系统重新加以利用。这里涉及报废产品回收政策和体系建设。回收后的报废产品必须能拆卸成零部件,并进行分类,由此产生“面向拆卸”的设计技术,零部件编码分类技术等概念全新的设计制造技术。报废产品解体后,面临一个如何利用的问题,从而出现各种翻新重用技术。

可持续发展思想在产品全生命周期中的贯彻,产生了“清洁生产”(Cleaner Production)模式。清洁生产是指将综合预防的环境策略持续地应用于生产过程和产品中,以便减少对人类和环境的风险性。对生产过程而言,清洁生产包括节约原材料和能源,淘汰有毒原材料并在全部排放物和废物离开生产过程以前即减少它们的数量和毒性。对产品而言,清洁生产策略旨在减少产品在整个生命周期过程(包括从原料提炼到产品的最终处置)中对人类和环境的影响。清洁生产包括:

1) 清洁的能源 如太阳能、水电、风力以及现有燃料的清洁燃烧。

2) 清洁的材料 低污染,无毒、易降解,可回收。

3) 清洁的制造过程 不消耗和污染环境资源,无三废排放。



4) 清洁的产品 使用中不污染环境, 报废后易回收。 程。

(3) 可持续发展的评价

第二再生产过程 社会物质资料再生产过程。

1) 三大再生产过程 (见图 36-2-6)

第三再生产过程 自然资源和生态环境的耗损与

第一再生产过程 人类自身的再生产与发展过 恢复过程。

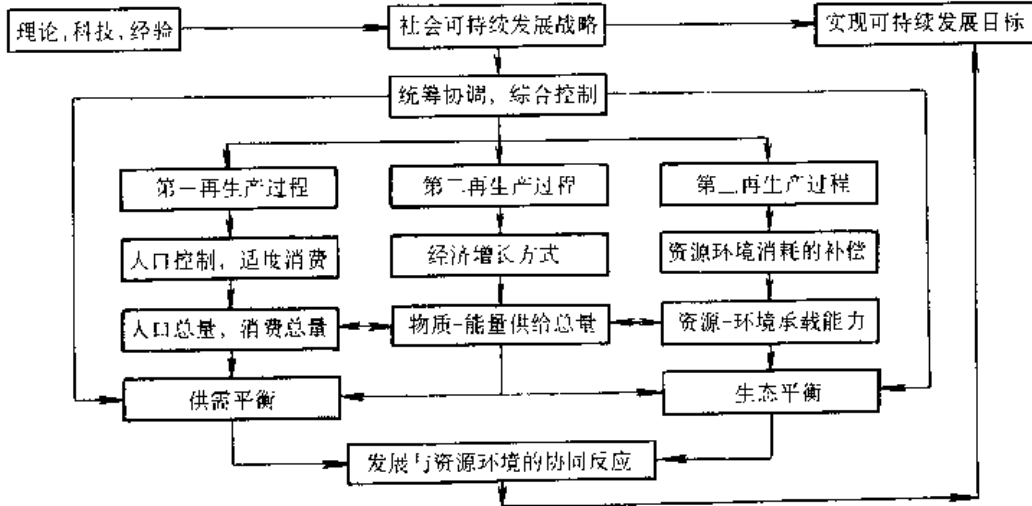


图 36.2-6 人与自然、发展与资源环境的协调机制系统框图

第一与第二再生产过程之间构成供需平衡关系, 第二与第三再生产过程之间构成生态平衡关系, 前者反映生产与经济发展, 后者反映环境支持能力, 两者之间应处于协同反应状态。这种协同性可从如下几方面来评定:

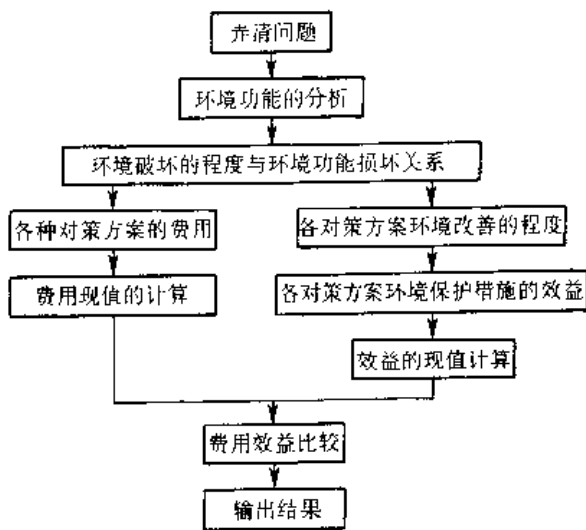
- 1) 三大再生产过程是否在可持续发展战略下处于有序状态;
- 2) 三大再生产过程是否形成正向的合力作用;
- 3) 为实现协调所作的投入与效益。图 36.2-7 为这项分析的过程结构。

### 参 考 文 献

- 1 王成森主编. 现代机械设计——思想与方法. 上海: 上海科学技术文献出版社, 1999
- 2 辛一行主编. 现代机械设备设计手册: 第 1 卷设计基础. 北京: 机械工业出版社, 1996
- 3 徐灏主编. 机械设计手册: 第 1 卷. 北京: 机械工业出版社, 1991
- 4 李康. 可持续发展的理论与宏观调控模式. 环境科学研究, 1997 (1) : 35

图 36.2-7 实现协同的投资与效益分析

2) 可持续发展的评价 可持续发展观的最基本点就是协调, 对可持续发展观念实施效果和影响的评价, 也应重点放在协调上, 即: 发展与资源环境和协调: 人与自然和谐。



# 第 37 篇 绿色产品设计

主 编 蔡建国  
编写人 蔡建国

## 第1章 概述

人口、资源和环境是困扰并威胁当今人类社会生存与发展的三大问题。因此,世界各国政府和人民都对此给予高度关注与重视。联合国环境规划署也为此先后召开过多次具有历史意义的重要国际性会议,遂使世界各国逐步对上述三大问题的严重后果形成了共识,并把可持续发展作为各国发展社会与经济的共同战略。于是,控制人口、节约资源和保护环境,便成为各国发展社会与经济时所考虑的首要问题。

### 1 绿色产品设计思想的由来

#### 1.1 人类社会生存发展面临的三大威胁

##### 1.1.1 人口问题

人口问题严重制约着人类社会与经济的发展。根据人口学家的预测,到本世纪50年代,全世界的人口将比现在的60亿翻一番,即达到120亿人。然而在未来的50年里,地球还是这么大一个地球。显然,要为新增加一倍的人口提供衣、食、住、行和就业机会,无疑会使地球更加不堪重负,会给未来的城市带来无穷的压力、会造成极大的社会问题和经济问题。为此,每一个国家,特别是发展中国家和经济落后而极度贫困的国家,都必须重视人口增长的问题,必须切实有效地实施旨在节制人口的生育(或家庭)计划。

##### 1.1.2 资源问题

滥用资源是导致地球不可再生资源严重短缺、乃至濒临枯竭的主要原因。制造业固然为人类创造了巨大的社会财富和灿烂的物质文明,但是也给人类赖以生存的环境带来极大的负担。下面的数据足以说明资源被滥用的严重情况,尤其是经济发达国家更为严重。

据联合国环境规划署的一份统计资料报道,目前世界上每消费1t产品,在其整个生命周期内,要给环境增添8t负担。我们把它形象地称之为“冰山现象”,如图37.1-1所示。图中说明:人们每天使用着各种形形色色的产品,这好比人们只看到浮出水面的冰山顶部那么一丁点儿;然而支持露出水面这一丁点儿的,却是沉在水下硕大无比的冰山底部。这顶部与底部的重量比例关系,即如图中所示的1:8。这里所指的环境负担,包括产品全生命周期内各阶段(设计、生产、流通、使用、维修直至报废处置)所消耗的资源 and 所排放

的废弃物二者的总量。这一数字着实惊人!如果我们对于这种滥用资源的状况继续放任而不加节制,则地球上的资源必将迅速消耗殆尽,而各种废弃物的迅速增长,也必将酿成公害而威胁人类自身的生存。

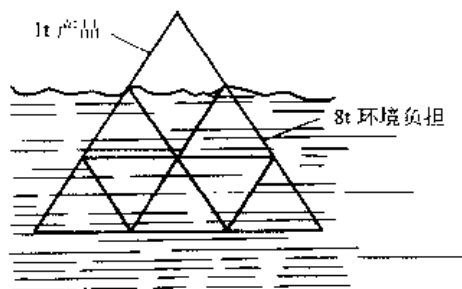


图 37.1-1 产品生产与环境负担间的冰山现象

现代社会的商品经济,实质上是一个物料密集型的经济。这可由图37.1-2中看出。图中表示产品的生命周期及其物料流。在产品生命周期的各个阶段,都有物料在流动。据统计,美国每年每人消耗的物料高达10t(不包括水、空气、石头等)。其中只有6%体现在耐用商品上,而其余的94%则都变成废弃物。最容易为人们所看见的影响环境的城市的固体废弃物只是产品生命周期生成的总的废弃物中的极小一部分。人们对废弃物在认识与现实之间存在的巨大反差,即如图37.1-3所示。图中清晰地描绘了从材料开采到加工、生产产品、配送、使用、直至最终报废/回收为止,每年产生的废弃物。从图中可以看出,城市固体废弃物仅占总的无害废弃物中的1.9%;而制造业所产生的废弃物则占总的无害废弃物中的60%。显然,产生的废弃物愈多,则其消耗的资源也愈多。因此,制造业理应在节约自然资源和防止环境污染中发挥主导作用。

在评价滥用资源方面,有人用生态脚印(Ecological Footprint)作为一种指标。所谓生态脚印,是指满足一个人的平均基本需要所折算的土地总量。图37.1-4便是美国、荷兰和印度各自生态脚印的比较。由图可知,满足一个美国人的平均基本需要,要12.2英亩;满足一个荷兰人的平均基本需要,要8英亩;而满足一个印度人的平均基本需要,却只要1英亩。因此,满足荷兰人的生态脚印,要求荷兰的面积为现在而积的15倍;而满足印度人的生态脚印,则要求印度的面积为现有面积的1.35倍。如果全世界的人口要像美国人那样生活,则需要3个地球方可维持。这从一个侧面反映了发达国家与发展中国家在生活水准上

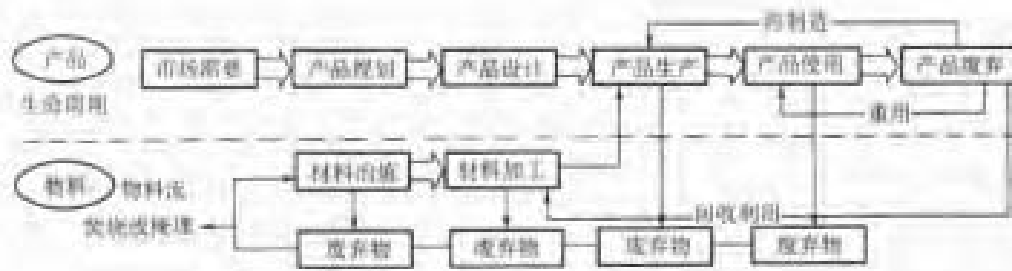


图 27-1-2 产品的生命周期及其物料流

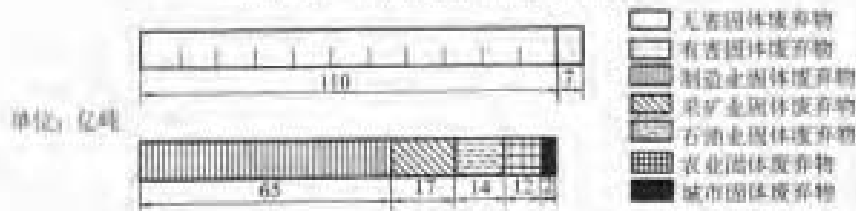


图 27-1-3 美国 RCRA 1988 年统计的固体废弃物



图 27-1-4 美国、荷兰和印度的生态脚印比较

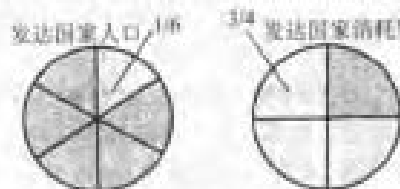


图 27-1-5 全球极不合理的资源消耗

的显著差异，从而也可看出两者在消耗资源上的巨大差距。

图 27-1-5 更进一步说明上述这种极不合理和极不公平的差距有多大。仅占世界人口 1/6 的北方发达国家，却消耗世界能源和资源的 3/4；而占世界人口 5/6 的南方发展中国家，则仅消耗世界能源和资源的 1/4。不仅如此，发达国家在能用资源的同时，还排放大量工业的、消费的、甚至有毒和有害的废弃物。

据专家估计，如果按照目前这样不加节制的资源消耗水平，用在本世纪中叶，像石油和煤，以及许多贵金属均将告罄。

发达国家是典型的利己主义者。第二次世界大战以前，它们奉行的是赤裸裸的炮舰政策，凭借它们强大的武力，它们到处疯狂地掠夺殖民地，强行将别国的资源据为己有。第二次世界大战以后，它们改弦易辙，不

再进行明目张胆的武力掠夺，而是改用更为隐蔽狡诈的经济侵略手段。于是，发达国家的跨国公司便纷纷在世界各地开设工厂，借此变相占有别国资源和劳动力来生产供它们享用的各种产品。这就是为什么占世界人口的 1/6，却可以享用世界资源 3/4 的缘故。

对于发展中国家而言，由于科学技术落后于西方，因此，国民经济生产总值每单位产值的资源消耗远高于发达国家约 3~4 倍。二次资源的利用率也仅为发达国家的 1/3~1/4。在平均能源利用率方面，美国为 50%，日本为 57%，而我国仅为 28%，由此可见，发展中国家存在着更为严重的资源浪费问题。

### 1.1.3 环境污染问题

在产品生命周期中，除了各个阶段消耗资源外，同时还排放出大量废弃物（如废气、废液、固体废弃物等），其中有些废弃物对生态环境和人体健康甚至是有毒有害的。正如上述联合国环境发展署的那份统计报告中所指出的，每生产 1t 产品，其所消耗的资源 and 排放的废弃物的总量可达到 8t。它既说明资源被滥用，而且也说明废弃物排放很严重。这大量废弃物若不经严格处置，而直接排放或任意丢弃，则必将污染环境（大气、土壤和水资源），以致引起严重的公害。

#### (1) 气候变暖的温室效应

全球气候变暖主要是由于人类生产与消费过程中所排放的废气（如 CO<sub>2</sub> 等）所引起的。这类废气阻碍了所释放的热量从地球表面和大气层中散去，从而形成了“温室效应”。据统计，大气中的 CO<sub>2</sub> 的浓度比工业化前提高了 25%，而且目前还在以每年 0.5% 的速度递增。英国的一项最新研究估计，按气象模型预测，到 2050 年时，冬季的平均气温将比现在提高 2.0~

2.5°C。人们已经感受到温室效应所引起的气候异常。

气温的提高还会导致南极和北极的冰山与冰川的融化,从而造成海平面的上升。国际上有一个名叫“政府间气候变化专家委员会”(IPCC)的组织,据它预测,到2100年,海平面将比现在上升0.3~1.0m,一种最佳的推测为0.5m。海平面如此上升,将直接威胁着沿海城市、岛屿和河流出海口地区的安全与生存。

### (2) 大气中臭氧层遭到破坏

来自太阳的紫外线,尤其是紫外线-B,会对生物圈中的各种生物造成伤害;而大气中的臭氧层则是阻碍紫外线进入大气的天然屏障。然而,现代人享用的冰箱、空调器中所使用的制冷剂—氟利昂,它所排放的氯氟烃类物质会消耗臭氧。因此,臭氧层遭到了破坏,使大气层出现空洞而增加了紫外线对地球表面的照射量。这是地球上人类白内障和皮肤癌等疾病发病率增加的重要原因之一。

### (3) 大气污染

大气污染包括光化学烟雾、悬浮微粒和酸雨等。

1) 燃料燃烧时所排放的氮氧化合物和挥发在大气中的汽油以及油漆溶剂等挥发性有机化合物(VOCs; Volatilized Organic Compounds),经过紫外线的作用发生反应而生成的一种光化学氧化剂。这种光化学氧化剂又称地面臭氧,是光化学烟雾的主要成分,具有极强的氧化作用。它几乎能与任何生物组织发生反应。因此,工业生产的不断发展和现代社会使用愈来愈多的各类汽车,使大气中的光化学烟雾含量日益增多,以致使人们吸入后引发严重的呼吸系统疾病,如:呼吸道发炎、肺功能损害引起胸痛、气短、咳嗽等。

2) 城市中大规模的基础设施建设和基本建设等的土方工程,交通工具在马路或公路上行驶时,都会扬起尘埃;火力发电设施和工业生产中所用燃料燃烧时,都会排放烟道废气;这些尘埃和废气中包含着大量的悬浮微粒。大气中总的悬浮微粒平均值高达432mg/m<sup>3</sup>,其中约30%是未经处理的。显然,人体吸入这种悬浮微粒将有害身体健康。

3) 随着工业生产的发展,煤炭以及其他化学燃料的消耗量在不断地增加,因此排入大气中的(SO<sub>2</sub>)、与(NO<sub>x</sub>)、等愈来愈多,以致使空中的雨、雪、雾、露等都呈酸性,一般统称为酸雨。如今,酸雨已经成为全球最严重的环境问题之一。工业上的燃料经燃烧后生成的烟雾,大都由高耸入云的烟囱排入空中。由于大气环流的作用,使之能飘扬过海而达千里之外,所以酸雨又被称作“跨国恶魔”。酸雨的主要危害是破坏森林生态系统,严重损害植被,改变土壤性质,抑制土壤中有机的分解,使土壤贫瘠,影响植物发育生长。同时它还破坏水生生态系统,使江河中大量水生动植物死亡。此

外,酸雨也使水源酸化而导致水中金属元素被析出,这对人体健康产生有害影响。

### (4) 水体污染

工业生产所排放的废液和废水中含有许多有毒有害物质,使水体严重污染,有些水体的水质不仅人类无法饮用,甚至令水中生物也无法生存。不少水体中的生物需氧量(VOD)值和氮与磷含量都严重超标。氮与磷是使水体富营养化的最重要的营养化元素。表层水的营养化,促使水面的藻类疯长,导致水下缺氧;而藻类腐烂时,又将消耗大量氧气,因此,水体富营养化将对水中的动物种群和植物种群的生存产生严重影响。

据报道,美国向海洋排放的工业废弃物约占全球的1/5,而仅废水一项每年高达200亿吨,其耗氧量为一般城市废水的3~4倍。这些废水中含有高浓度的氰化物、酚等剧毒物质,以及铅、镉、铜、汞等金属和总量为100多万居里的放射性物质。由于近海污染严重,使80%的贝类无法食用,有些鱼类的含汞量也很高。由此可见,水体污染已经危及海洋渔业的生存和发展。

必须指出的是,发达国家不仅如上所述地消耗掉世界上75%的资源,享受着由这些资源所创造出来的物质财富,与此同时还将排放出大量废弃物来污染环境。特别像发达国家的汽车,拥有量占了全球的绝大部分,因此,它们所排放出来的汽车废气便成为全球气候转暖的主要原因。更有甚者,由于发达国家制订了一系列严厉的保护本国环境的法令和法规,因而许多排污严重的企业便无法在本国生存;为此,它们将这类企业迁至发展中国家。这样既可利用当地廉价的资源和劳动力,赚取巨额利润,而且又把污染转嫁于人,以实现它们损人利己的可耻策略。

以上只是列举了与工业生产有关的滥用资源和环境污染问题的一些主要方面,从中已能看出问题的严重性。面对这些严重威胁人类生存的问题,从20世纪60年代末开始,人们便萌发出强烈的保护环境意识。首先是西方国家人民群众的觉醒,掀起了波澜壮阔的以保护环境为宗旨的绿色运动。从此,声势浩大的绿色运动,便此起彼伏,遍及全球。

对于向社会提供产品和服务的制造业而言,理应在贯彻“可持续发展”战略、实施“节约资源和防止污染”方面发挥重要作用。因为在产品制造阶段时要消耗资源,在产品包装运输阶段时也要消耗资源,在产品使用和维修阶段时还要消耗资源,所以这就是使制造业成为资源消耗大户的缘由。显然,在各个阶段消耗资源的同时,必将向环境排放出各种废弃物。形形色色的工业废弃物,则正是危害环境的主要污染源。既然消耗资源与产生污染都与产品生命周期的各个阶段有关,因此,要想节约资源和防止污染,便需从产品入手,从产

品全生命周期入手。于是,便提出一种新的产品设计理念。这便是本篇所要介绍的绿色产品设计。

## 1.2 绿色产品设计的发展概况

### 1.2.1 国外的发展概况

20 世纪 60 年代末以来,针对多次发生的化学品生产厂家的重大事故和核电站可怕的放射性泄漏事故,以及由此造成的严重威胁人民生命健康的公害,激起了国外高涨的绿色群众运动,人们禁不住要发出保护人类生存环境的强烈呼声。这也不得不引起国外政界、学术界和工业界的高度重视。正是在这种环境情况日益恶化下:

- 国外的政府部门积极制订了各种旨在保护环境的相应政策、法令和法规,进一步健全市场机制,大力推行环境(生态)标志;
- 国外的学术界高度重视节约资源和防止污染以减轻环境负担的严重问题,因而展开了有关绿色产品、绿色设计、绿色制造、清洁生产、环境标准等方面的研究,提出了相关的新概念、新理论、新技术和新方法;
- 国外的工业界则在政府有关政策、法令、法规的指引和约束下,以及在公众和舆论的监督下,开始正视并正确处理企业自身的持续发展与其责无旁贷的治理环境之间的关系,并在企业的产品开发和生产过程中,积极采用各种研究开发成功的、旨在减轻环境负担的新技术和新方法。

### 1.2.2 国内的发展概况

在保护环境的立法、环境参数的评价标准、排污的处理技术、绿色设计的方法等方面,我国都远远落后于国外。正因为立法和标准尚不健全,技术和方法还不完备,企业对于开发设计绿色产品也未引起足够重视,政府环保部门则主要着眼于大环境(水源、土壤和空气的质量)的治理,所以国内在绿色产品设计的研究与应用方面,只能说是处在刚刚起步的阶段。目前只有一部分高等院校和科研机构在进行绿色产品设计理论和方法的研究,还有部分企业则通过产学研的形式,正在进行废旧产品回收、拆卸、再制造的试点,或是从事废弃物的无毒无害处理技术的研究。

为了在制造业中贯彻可持续发展的战略,目前国内的企业主要限于孤立地从产品设计这一个阶段,而非从产品全生命周期的各个阶段,去考虑节约资源和防止污染的问题。因此,只是在产品设计这一阶段做节约产品本身制造时所需资源的文章,是不可能有着显著的实际效果的。这时所能取得的效果,充其量,也仅

仅是在产品设计时采取了下列已有的措施所致:

- 产品小型化 设计产品时务必使产品的总体(外观)尺寸尽量缩小;
- 产品轻量化 设计产品时力求尽量减轻产品的重量。

通过实施产品的小型化和轻量化,使产品变得小而轻,这样确实能使产品既节约能源,而又节省原材料。由此可见,国内企业目前主要通过在设计阶段实施“节能省料”方针,初步达到节约资源的目的。

至于防止污染,由于立法不够健全,尚不具备约束力,加之设计人员的环境意识亟待提高,因此目前大部分产品在其设计阶段时,尚未充分考虑防止环境污染的问题。只有像冰箱、空调器、内燃机、汽车等与国际接轨的产品,因为已有禁止使用氟利昂作为制冷剂的“蒙特利尔议定书”和欧洲制订的内燃机与汽车尾气排放标准的约束,所以在设计这类产品时考虑防止在使用阶段的污染问题。

我国加入世界贸易组织(WTO)后,意味着我国企业可以参与全球化的市场竞争。我国企业要将自己的产品打入国际市场,便必须首先使自己的产品符合国际环境标准。尤其是极端利己的发达国家,为了保护自身国家的环境不受污染,它们一方面将有毒有害的污染工业扩展到发展中国家;另一方面又制订了严格的环境标准和环境立法,借以构筑所谓的“绿色壁垒”,以阻挡和遏制不符合它们环境标准和环境立法的外来产品的输入。

由此可知,为了保护我国自身的环境,我们也必须一方面迅速制订并健全与国际接轨的我国环境标准和环境立法;另一方面,也应使我国企业深切地明确实施绿色产品设计的必要,否则,我国企业的产品便无法进入国际市场。比如,我国的汽车要出口欧洲,我国汽车的尾气排放,就必须达到欧洲规定的标准;也即我们必须按欧洲的尾气排放标准,来设计我们的汽车,否则人家便搬出绿色壁垒而不予准入。

## 1.3 实施绿色产品设计存在的问题

此处主要从产品的生产者和产品的使用者,这两个方面来分析。

### 1.3.1 产品生产者存在的问题

许多企业把自身的持续发展与其理应治理环境的责任对立起来,即存在如上面所提到的、江泽民同志严厉批评的“先污染,后治理”的错误思想。产生这种思想的原因,就在于企业过分看重眼前利益,认为治理环境将使企业增加投资,也即增加企业的支出,从而增加产品的成本,因而减少企业的利润。为此,许多企业宁

肯把治理环境所追加的投资用于发展生产,这样可使企业直接得到实惠,可以看见自身的持续发展。于是,它们便把自身的持续发展与治理环境对立起来。显然,这种认识是片面和错误的。

### 1.3.1.1 绿色产品设计能为企业带来效益

实际上,通过实施绿色产品设计,能为企业带来极大的好处。这可从下列两方面看:

- 由于运用绿色产品的思想,促使企业重视资源的利用率,因此,可以
  - 降低成本(节约了资源);
  - 缩短生产时间(减少了加工裕量)。
- 因为绿色产品设计贯彻的是产品全生命周期的设计思想,所以它必然要把与产品全生命周期各个阶段有关的职能部门人员组织起来,形成一个集思广益的团队,并且按照并行工程的方式开展工作。这样便可有效地
  - 促进产品结构的创新;
  - 促进产品工艺的革新。

### 1.3.1.2 绿色产品设计产生巨大效益的实例

这里将列举国外著名企业成功应用绿色产品设计而取得巨大效益的一些实例,以说明企业自身的持续发展与治理环境的责任,是彼此相辅相成、相得益彰的,而并非彼此对立、互相矛盾的。

#### ■ 舒乐(Xerox)复印机公司

舒乐公司制订了其产品无废弃物的环境目标。为了实施舒乐的这一战略目标,将其产品结构的设计成

- 整机可以再制造(Remufacturing);
- 零件可以重用(Reuse)。

为此,舒乐公司在产品开发与设计时,建立了以下绿色设计理念:

- 产品必须易拆;
- 产品及其零部件必须具有耐久性(长寿命);
- 产品及其零部件必须能再利用;
- 在产品时必须考虑其回收;
- 应使产品及其零部件的潜在寿命终止期(End-of-life)为最大。

如今,舒乐公司所设计的产品中,90%是可以再制造的。通过产品回收后的拆卸、清洗、检测、再制造,零部件便可在新产品中得到重用。如此,舒乐公司在贯彻其资产回收管理规划(ARM: Asset Recycle Management)中,每年节约额可高达数亿美元。

#### ■ 伊莱克斯(Electrolux)家用电器公司

伊莱克斯是欧洲一家著名的家用电器的制造商。家用电器是耗电较多,而又排放有害废弃物(如空调、冰箱等)的一类产品。因此,严格按照环境参数来开发和设计产品,则显得尤为重要。伊莱克斯为此制订了自

己的环境战略和环境参数,并在产品开发与设计付诸实施,从而取得了喜人的成就。

根据1999年伊莱克斯公司的报告,公司的白色产品(指冰箱等家用电器)在实施公司的环境战略后,使产品具有最佳的环境性能,从而使其产品在欧洲市场的销售额达到21%,利润达到31%。这些数据表明:重视环境治理工作,能给企业带来可赢利性。

#### ■ 3M公司

人们十分关注3M公司用于录音带和录像带等产品的包装材料,诸如纸板箱、泡沫塑料以及收缩性薄膜等,在它们被废弃时对环境所造成的有害影响。为了最大限度地减少公众所关注的环境影响,3M公司的设计人员,专门研究开发了一种可以回收而重用的高密度聚乙烯(HDPE: High Density Polyethylene)制成的新型包装箱。这种新型包装箱既灵活,又易于堆放,而且还能最少使用40次。

由于采用了这种新型包装箱系统,不仅节省大量包装材料,而且还能节约达400万美元。

这个例子充分说明了,即使仅仅在产品全生命周期的包装环节上考虑绿色设计,也能为企业带来极大的经济效益。

基于绿色设计思想,还开发了新一代的塑料擦洗垫。新的擦洗垫要比上一代的产品好,而且具有对环境友好的许多优点。这就满足了顾客对清洗的需求,因而赢得了顾客的赞赏和好感。这是由于

- 3M公司采用了新的材料聚乙烯酯(PET)来制造新一代的擦洗垫,这样可以减少对原生材料的需要。
- 擦洗垫中所用的肥皂可以生物降解,并且不含磷。这就使3M的擦洗垫比一般的钢丝绒更耐用。这种擦洗垫可多次重新注入肥皂,从而使其有第二次生命周期和第三次生命周期,延长了擦洗垫的使用寿命。

新的含肥皂的3M塑料擦洗垫,已经占有13%的市场份额,并且已在无擦伤的擦洗垫市场中独领风骚。

这个例子则说明按照绿色设计的要求,在开发设计新产品时,正确选择和研制材料,同样可以为企业节约原材料,降低成本,延长产品寿命,提高顾客的满意度,增强企业的市场竞争力,给企业带来可观的经济效益。

#### ■ 戴尔(Dell)计算机公司

戴尔公司是美国一家闻名遐迩的个人计算机制造公司。戴尔公司向顾客提供了崭新的产品租赁和资产恢复服务。这样便解除了最终用户对自己手中过时计算机的处置负担。戴尔公司对所生产的计算机,完全采用绿色设计理念,所以戴尔公司的计算机是按可维修、可拆卸和可重用的思想来设计的,因而公司便可将过去租赁而回收的计算机(其产权为属于公司所有的资产)重新修复翻新(也即所谓的资产恢复)上市,这样

便延长了计算机的寿命,从而避免将计算机丢弃和掩埋,同时也降低了重新上市的计算机的制造成本。

#### ■ 丹麦铁路 (Danish Railway) 公司

绿色产品设计虽能导致产品改善对环境的影响,但是这并非仅限于对单独一种产品而已,有时也会影响整个运作系统。这里所举的这个例子,便属于这种情况。

丹麦铁路公司开发的一种 S-型列车,是根据一系列经过仔细研究的环境准则来设计的。这些准则归纳为:

- 材料选择——尽量优化材料的使用;
- 零件必须可重用;
- 保证在运输系统整个生命周期能有效利用能源。

在 S-型列车的绿色设计中,可以看出,不仅列车本身,而且整个运输系统,都得益非浅。其中所取得的成就大致有:

- 运载能力增加了 35%,而整车的重量却减轻了 46%;
- 列车轻量化后,它的启动和停止更快,因而缩短了列车的运行时间;
- 轻量化后,降低对能源的需求约 10%~14%;
- 列车是由发动机驱动的。当制动时,发动机使权且充作发电机使用,因此,部分被消耗的能量又不断回送给系统,借此使系统能节约能源。

从以上所列举的这些实例中,足以说明,企业自身的成长和发展是与积极重视环境治理不矛盾的。上述那些国外知名的公司,由于采用了绿色产品设计,不仅使产品具有更好的环境属性,节约了资源,防止了污染,而且还提高了顾客的满意度和企业自身的形象。当然,与此同时也给这些企业带来了丰厚的利润。

### 1.3.2 产品的使用(消费)者存在的问题

传统的消费者往往还不习惯于使用再生的或重新上市的产品,这也是影响实施绿色产品设计的一种障碍。其原因就在于,绿色产品设计必须考虑产品生命终止后的处置问题,也即生命终止后的产品必须回收,必须使回收后产品中的零部件,能够最大限度地得到重用,或者重新回炉再生成二次材料。这样,在重新上市的产品中,便包含了许多被重用的零部件和用二次材料制成的零部件。这种已被使用过一个生命周期的零部件是否依旧可靠?将必然成为顾客心目中一个挥抹不去的疑虑。顾客当然不愿意花钱去买一件貌似新的,而实际上却已被别人使用过一茬的“产品”。顾客的这种心理,应该说是可以理解的。

然而,顾客的这种心理,却给一部分不重视减轻环

境负担的企业援以借口,使它们犹豫是否还有继续实施绿色产品设计的必要?因为假若如此设计的产品很少有人问津,则企业努力生产绿色产品的目的和意义又何在呢?

#### ■ 通过宣传教育转变消费者的思想认识

为此,整个社会必须努力通过各种方式的宣传教育,以转变顾客的传统消费观念,使之愿意接受再生产品。这样,才能拓宽绿色产品的市场,才能推动更多企业更好地实施绿色产品设计。例如:国外最初在推广使用再生纸时,受到很大的阻力。不少人认为色泽不好,而拒绝使用;但是,当意识到使用再生纸能极大地节约木材资源、保存大片森林时,人们便逐渐接受了它。如今,不仅一般的书写纸是再生的,许多国外著名的贺卡制造商,也都采用再生纸印制价格低廉却又十分精美的贺卡,甚至在美国民航客运班机上所提供的餐巾纸,也标明是用再生纸生产的。这也从一个方面,说明转变观念、提高人们节约资源和保护环境的意识是何等重要!

#### ■ 改变传统的商品销售方式为租赁方式

除了宣传教育外,改变传统的商品销售方式为租赁方式,也不失为是一种使消费者乐意接受再生商品的有效途径。传统的销售方式,实际上是一种消费者买断商品产权的方式。显然,如上所述,要让消费者掏大钱去买断再生商品的产权,他是不会心甘情愿的。他心里总感到占有这样的商品不舒服。如果采用租赁的办法,消费者只需花很少的租金,就可享用商品的使用权,他何乐而不为呢?其实,就消费者而言,消费者只需要产品的使用权,他并不需要买断产品的产权。因此,在租赁制下,商品的产权是属于生产商的。仅仅改变商品的交易方式,就能给“可持续发展”和绿色产品带来如下所述的莫大好处。

#### ● 企业自动实施产品收回制度

由于产品的产权是属于企业的,所以当产品生命终止后,企业便责无旁贷地要收回属于自己的产品。这样便解决了一个目前令人头痛的废旧产品难于回收的老大难问题。只有把企业自己生产的产品收回来,才有可能考虑零部件的重用问题。在国外,许多国家已经通过立法,制订了“谁生产,谁收回”的法律(Product Take Back Law)。从而确立了产品收回制度。使企业明确收回产品是自己的责任。实际上,在前面列举的舒乐公司和戴尔公司的例子中已经提到,舒乐公司生产的复印机和戴尔公司生产的个人计算机,在欧美等国的市场,都是采用租赁方式供给消费者使用的。

#### ● 解决就业岗位

废旧产品收回后,便需对其进行分解、拆卸、清洗、检测、再制造。完成这些作业,就需要大量的劳动力,



因此可为社会解决大量就业岗位。这是由于租赁制而带来的巨大社会效益。在我国,则更有其现实意义。许多下岗职工原本就是从事制造行业的工作,如今为了就业,往往要花费一定代价接受重新培训,学习一门新行当、新手艺。然而,参加废旧产品收回的工作后,他们却可以检回自己的老本行,驾轻就熟地发挥自己的技能和专长。

#### ● 真正做到“物尽其用”

物尽其用,其实也是一种有效的节约资源的方式。由于目前通行的产品产权买断的交易行为,而买断后的产品却未必就能被充分利用,因此许多商品的利用率是极低的。尽管国外的轿车保有量很高,但是其利用率却很低。每天大部分时间是闲置在车库或停车场。这不仅使轿车未能做到物尽其用,而且还占用大量宝贵的土地去修建车库和停车场。为了改善这种情况,国外也正在研究数人共享一辆轿车的各种分时-调度方案。目的就是要物尽其用,杜绝闲置不用这另一种形式的资源浪费。商品的租赁制,至少能满足这一目的。

### 1.4 节约资源和防止污染的策略

为了节约资源和防止污染,目前在制造业中有两种研究绿色产品设计的思路。

#### 1.4.1 “源头预防”策略

从产品设计一端(产品生命周期的源头)着手研究绿色设计问题,这便是所谓的“源头预防”策略。它是指在产品设计阶段,便从产品全生命周期的角度去预防滥用资源,并尽可能地防止产生污染。这样,便较为直接地保证通过产品设计过程,使所设计的产品具有良好的绿色度。如图 37.1-6 所示,即为一种从产品生命周期的始端研究产品绿色设计的模型。

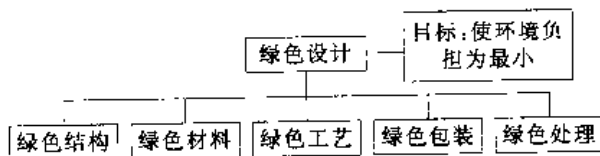


图 37.1-6 一种产品绿色设计的模型示

#### 1.4.2 “末端治理”策略

这一策略从产品回收一端(产品生命周期的末端)入手研究绿色设计问题,所以称作“末端治理”策略。它是指在产品生命周期终止(寿终正寝)后,要对报废产品进行有效处置。这包括废弃产品的回收、拆卸、清洗、检测、再生(翻新、修复、回炉)、重用、焚烧、掩埋等。通过对废旧产品的回收拆卸过程,从中找出原设计存在的不符合绿色要求的问题与缺陷,然后总结归纳成设计准则反馈给产品设计部门,以便在重新设计或改进设计时能按这些准则进行。

从长远来看,这两种策略既不应对立,也不可偏废,而应该彼此相辅相成。就大部分现有产品而言,由于这些产品在当时设计之初,可持续发展的战略尚未提出,设计人员很难从可持续发展的战略高度去考虑,因此无法在产品设计一开始便贯彻“源头预防”的策略,以致使这些产品对环境的友好性很差。正因为如此,所以对于这些现有产品,只能采取“末端治理”的策略。

显而易见,“末端治理”的策略,是一种消极被动的事后处置策略;而“源头预防”的策略,则是一种积极主动的事前防范策略。从倡导绿色产品设计角度来看,必须主张“源头预防”的策略,也即在产品研究、开发和设计之初,便充分运用产品全生命周期的观点来考虑节约资源和防止污染,如此方可使产品在其全生命周期内对环境所造成的负担最小。事实上,只有在产品设计阶段切实贯彻可持续发展的战略,才能最有效地节约资源和防止污染。

即使在产品设计阶段采取了“源头预防”的策略,这也并非表示就无需“末端治理”了。其实,只要产品报废,也即产品到了它生命周期的末端,就必须接受末端治理。做好“源头预防”,可以使“末端治理”能够更容易、更合理地进行。反过来说,“末端治理”也并非真的消极被动到了无所作为的地步。事实上,通过“末端”对报废产品的拆卸,可以发现许多不尽合理的结构设计及材料选择。由此即可提出改进的建议,并总结成相应的设计准则,反馈给“源头”——即产品的研发和设计部门,以提高未来产品结构设计及材料选择的质量,同时也有利于今后的“末端治理”。此即前面提到的两者相辅相成的涵义。

## 第 2 章 绿色产品设计

### 2.1 概述

由于绿色产品设计是一个新的概念,因此无论在国外或在国内,人们对于这一新概念有着各种不同的称呼。例如,有人称为“绿色设计”(Green Design),有人叫作“生态设计”,(Eco Design)还有人(主要在美国)称之为“面向环境的设计”(Design for Environment),也有人根据产品生命周期不同阶段所进行的绿色设计,而分别称为“面向包装的设计”(Design for Package)、“面向使用的设计”(Design for Usage)、“面向维修的设计”(Design for Service)、“面向拆卸的设计”(Design for Disassembling)、“面向回收的设计”(Design for Recycling)等,真可谓不一而足。这也是很正常的。因为当一种新生事物刚出现时,人们对它还缺乏全面的认识和理解,所以往往会仁者见仁、智者见智。经过一段时间的成长、发展、成熟以后,随着人们对它的认识和理解不断加深而日趋全面,这时,对于这一新生事物的命名便会逐渐趋于统一。

### 2.2 传统设计与绿色设计

当人们日益关注所生存的环境质量时,人们期待和乐于购买的是绿色产品。于是便提出了绿色产品设计,希望通过绿色产品设计能设计出绿色产品。既然提出了绿色产品设计,人们不禁要问:传统的产品设计与绿色产品设计究竟它们的异同何在?是否两者截然不同?下面便就此问题进行分析 and 阐述。

#### 2.2.1 传统产品设计

##### (1) 传统产品设计的实质

传统产品设计主要考虑的是实现顾客需求的产品基本属性,即产品的功能、质量和生产成本。而忽略了产品的环境属性,也即产品生命周期各个阶段的环境影响。此外,传统产品设计并未考虑产品使用阶段的使用成本、保修期后的维修成本,乃至产品生命终止后的处置成本。

##### (2) 传统产品设计的方式

传统的产品设计方式,由于沿袭“设计是龙头”的思想,因此把产品设计局限在设计一个部门。由于产品设计部门所考虑的主要内容,仅限于产品的基本属性,所以它沿用传统的串行工程方式。如图 37.2-1 所示。

在传统的产品设计过程中,设计人员除要解决与

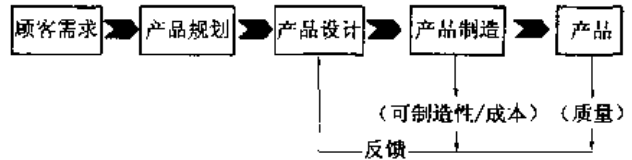


图 37.2-1 传统的产品设计方式

产品功能有关的产品总体结构、材料选用、零部件形状及尺寸以及零部件的联接形式外,在设计时还要考虑产品的质量、可制造性与成本等问题,但是这些问题究竟解决得如何?往往无法在设计阶段得到确切结果,而是有待于通过实践的检验。正如图 37.2-1 中所示,通过产品制造,可以检验产品是否具有有良好的可制造性,是否达到了预期的目标成本。也只有产品制成后,才能检验其是否符合规定的质量要求。如果通过产品制造发现可制造性、成本和质量都未能满足预期目的,则需将发现的问题和改进的意见反馈给产品设计部门,以便最终经过修正而能使产品结构符合设计意图。

虽然此处设计人员考虑了可制造性,可是传统上,这里所指的可制造性,仅限于与制造阶段本身有关的,而未考虑与使用、维修和回收等阶段有关的可制造性。至于成本,也只是考虑了制造成本而已,同样未计及产品的使用成本、维修成本和产品生命终止后的处置成本。这就是传统产品设计忽视了环境影响的缘故,因而未能从产品全生命周期的角度去设计产品。

#### 2.2.2 绿色产品设计

##### (1) 绿色产品设计的实质

绿色产品设计从产品全生命周期出发,考虑了生命周期各个阶段对环境的影响。特别是在产品生命终止后,绿色产品设计必须考虑废弃产品如何回收、再生和利用。只有这样,才能使废弃产品对环境的影响减至最小,才能最大限度地提高资源的循环利用率。

传统产品设计条件下,产品的生命周期是开环式的,如图 37.2-2 所示。这说明产品经设计、制造而进入流通领域,并交付顾客使用;随后在使用中经过多次维护与修理,最终因无法修复而告废弃。至此,产品生命終了。从图中可以看出,这时产品的生命周期呈开环形式。

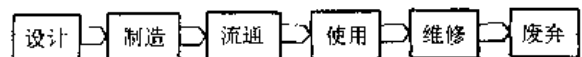


图 37.2-2 开环式产品生命周期

绿色产品设计条件下,产品的生命周期是闭环式的,如图 37.2-3 所示。这时,产品同样经设计、制造而进入流通领域,并交付顾客使用;随后在使用中也经过多次维护与修理,直至无法修复而告废弃。然而从图

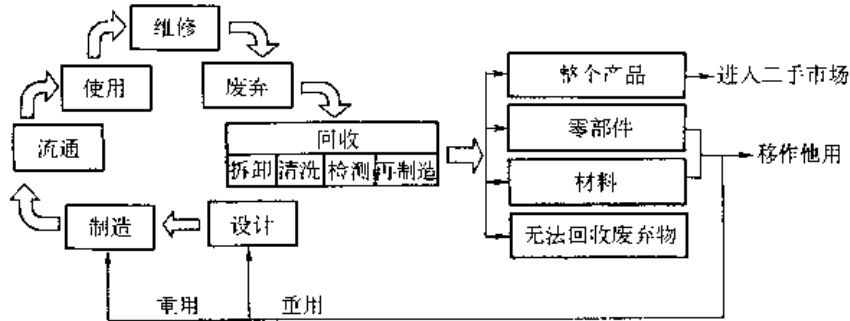


图 37.2-3 闭环式的产品生命周期

- 优先考虑（上策）：若整个产品经过稍加整修，则可作为二手产品进入二手市场。
- 其次考虑（中策）：若整个产品无法回收重用，则将其中的零部件通过再制造而重用于新产品或老产品中。此外，还可移作他用。
- 最后考虑（下策）：若零部件回收重用也不合算，则只能作材料回收。此时可将整个产品或零部件粉碎、分选、压块、打包，重新回炉冶炼成二次材料。
- 对于产品中无法回收利用的废弃物，则可考虑焚烧（尚可回收热能）或掩埋。

一般上述这类无法回收利用的废弃物，占产品所用原材料的 5% 左右。由此可见，产品中的原材料有 95%，是可以通过回收而得到重新利用的。仅仅只有 5% 的原材料是被废弃不用了。这也就是当前所倡导的循环型经济真谛之所在。

### (2) 绿色产品设计的方式

因为绿色产品设计要把产品生命周期各个阶段的环境影响，集中在设计阶段去考虑和解决，所以它必须采用并行工程的方式，以便把与产品生命周期各个阶段有关的人员，组织成设计团队。这样才能集思广益地进行绿色产品的设计。也正因为这样，通过并行工程方式，就能够把后续各阶段可能产生的问题，最大限度地 在产品 设计阶段予以解决。从而有效地防止了生产阶段不必要的工程更改，以免造成经济损失或延误交货期；同时也使日后废弃产品的回收处置，可以更容易和更经济。

目前，为了便于把产品生命周期各个阶段的环境影响考虑在绿色产品设计中，最简单的和最常用的做法，是把各个阶段的环境影响因素编制成检查表式的设计准则或设计指南，供设计人员在设计时参考和自我检查。这些设计准则或设计指南，将在第 3 章中介绍。

中可以看出，产品废弃后并非无人过问，而是通过回收，把可以重用的零部件和材料重新回归原来的生产系统实现循环利用。

废弃产品经过回收处置，可以分为 4 个层次：

绍。

### 2.3 传统产品设计与绿色产品设计的异同

综上所述，可以看出，绿色产品设计同样要考虑产品的功能、质量、可制造性和成本等产品的基本属性，同样要解决产品的总体结构、材料选择、零件结构形状和尺寸以及零部件的联接形式等基本设计内容。就此而言，绿色产品设计与传统产品设计的异同。

除此而外，传统产品设计所考虑的问题，局限于设计和制造出满足顾客需求的产品基本属性，至于把产品交付顾客使用后的环境负担问题，则不在其考虑范围之内。然而，绿色产品设计却要把这些问题纳入其设计范围以内。正如上面所说的，绿色产品设计不仅要考虑产品出厂后交付顾客使用和维修阶段的环境负担，甚至还要考虑产品废弃后的回收和处置（拆卸、清洗、检测和再制造）阶段的环境负担。这是绿色产品设计与传统产品设计的相异处。

由此可知，绿色产品设计并非是一种与传统产品设计完全对立的新的产品设计方法，它只是在传统产品设计的基础上，增加了对环境影响的考虑。因此，传统产品设计中的各种行之有效的的设计方法，包括各种计算机辅助设计技术，仍然能够广泛地用于绿色产品设计中；只是除此而外，还应将产品必须具备的环境属性，一并 在产品 设计阶段考虑在内。

### 2.4 绿色产品设计的基本概念与方法

20 世纪 80 年代末 90 年代初，随着填埋、堆放场地的日益紧张，环境问题的解决从“末端治理”走向“源头预防”，国外开始了产品全生命周期设计、面向环境的产品设计、面向拆卸设计、面向回收设计等方面的研究。许多学者和研究人员针对产品的可回收性提出了各自的理论。从 90 年代初丹麦技术大学 Alting 教

授最早提出生命周期设计的概念、倡导绿色产品开始,绿色产品设计(国外称之为面向环境的产品设计—Design for Environment)的研究就逐渐展开。目前美国、日本、澳大利亚和西欧各国(德国、英国、荷兰、比利时、意大利、瑞士等),在相关领域中研究力量的投入力度颇大。经过近 10 年的努力,国外在这一领域内已取得了值得瞩目的成果。

我国的经济近年来持续增长,工业技术水平也在不断提高。然而我国环境状况的恶化却成为令人深深担忧的问题。要走可持续发展的道路,就必须节约并充分有效地利用现有资源,保护环境。与其他国家相比,我国的资源消耗量高,二次资源利用率低,有相当一部分资源变成了污染物。如上所述,我国每单位国民生产总值所消耗的矿物原料比发达国家高 2~4 倍,而二次资源的利用率只相当于世界先进水平的 1/4~1/3。因此,对我国环境问题进行源头预防与末端治理,进行绿色产品设计的研究就更加迫切与重要了。我国近年来也开始在绿色产品设计领域开展一定的研究工作,并取得了初步的成果。作者曾经参与承担国家自然科学基金重点项目“机电产品绿色设计理论与方法”的研究,现在就研究所获成果和体会,对绿色产品设计的有关概念、原理和方法进行介绍。

### 2.4.1 基本概念

#### (1) 产品的生命周期(Product Life-cycle)

产品的生命周期可以分为商业生命周期和物理生命周期。产品的商业生命周期,是指产品从投放市场开始,然后经历成长期、成熟期、饱和期和衰退期,到退出市场而产生下一代产品的一系列商业过程。而产品的物理生命周期是按照材料与能量的流动来定义的。通常所谈论的产品生命周期是指它的物理生命周期。

产品的物理生命周期(见图 37.2-4)开始于材料的获取加工,经历了制造、运输和使用后被淘汰。淘汰的产品可直接回收重用(二手货);或经过拆卸、翻新可重用的零部件使其返回产品的制造过程;而对于无重用价值的部分进行材料回收使其返回材料加工;还有一部分材料被分解,作为燃料回收能量或产生再生原料用以材料加工,最后的废弃物被处置,回到自然界而形成一个封闭的周期。在产品的生命周期研究中,出现了“逆向制造”(Demanufacturing)的概念。逆向制造与制造过程进程方向相反,它是基于材料和产品回收重用的制造过程。

#### ■ 绿色产品设计(Green Product Design)

绿色产品设计作为一种系统的方法,在新产品的开发阶段就考虑产品在生命周期内对环境的影响,从而减少对环境的污染、对资源的浪费、对安全和健康等

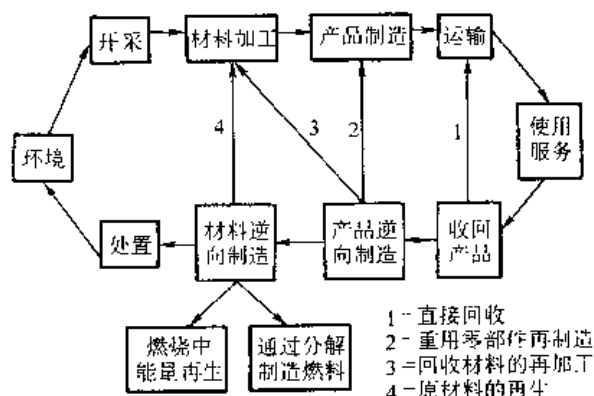


图 37.2-4 产品的生命周期

可能产生的负面作用。对制造业来讲,绿色产品设计意味着选择适当的材料,进行合理的功能结构设计、造型设计,制定恰当的制造装配工艺,从而设计出面向环境的产品。在并行工程中,绿色产品设计与其他的许多的设计要求一样,将产品性能(可测试性、可制造性、可装配性、可维护性等)设计贯彻于产品生命周期的各个阶段。

绿色产品设计的发展过程可分为 4 个阶段:①认识问题,形成设计指南;②问题清晰化,形成评价方法体系;③量化,应用于某些产品的设计过程中;④成熟,集成于设计过程与方法中,并应用于所有的产品设计。目前,绿色产品设计处于第二个发展阶段,出现了一些定性定量的评价方法,但这些方法还有待于进一步发展完善。在美国,绿色产品设计是由环保局、工业界、政府和科研机构联合发起的庞大的合作项目。它包括许多子项目,例如绿色电子产品设计(DFE Electronics)、绿色印刷产品设计项目(DFE Printing Projects)、自动表面修饰项目(Auto-refinishing Project)、绿色化学计划(Green Chemistry Program)、环境会计(Environmental Accounting)等。欧洲和日本等国家也对此有很大的投入。中国国家自然科学基金委员会近年来也开展并资助了一批有关的项目,但目前国内在这方面的研究还处于起步阶段。如果对现有的绿色产品设计方法、工具进行分类归纳,总体来讲,可把它们分为两大类:一类是基于产品生命周期评价(LCA; Life-Cycle Assessment)的生命周期设计(LCD; Life-Cycle Design);另一类是基于改进产品某一方面特性的面向 X 设计(DFX; Design for X)的方法的集成。在进一步讨论这两类方法之前,首先介绍一下有关的分析和评价方法。

### 2.4.2 绿色产品设计的分析、评价方法

对绿色产品设计而言,分析、评价方法是对产品的设计方案和 design 参数进行优化和筛选的关键,它决定

了设计者将如何进行决策。这里简要讨论一下通常使用的一些分析、评价方法,包括定性和定量的方法。

### (1) 检查表 (checklist)

检查表是最简单而且最广泛使用的一种定性评价方法。它把所要遵循的一些原则通过问题的形式罗列出来,设计者通过对这些问题的逐个回答和分析而不断地改进设计方案。通常,检查表被用作设计初始阶段的一种分析方法,因为检查表分析法对数据和其他一些参数的需要程度不高,而且便于维护,可以不断地更新问题以改进检查表,同时它又容易理解和完成。例如,检查表被广泛应用于材料的选择,它把符合环境要求的各种推荐材料以及对环境有不良影响的或禁用的材料罗列出来,以供设计者进行选择和优化。尽管检查表具有便于使用的优点,它也有很多不足之处。例如:

- 虽然检查表可以进行一些数量评分,但它还是一种定性的分析方法。也就是说它只能对产品性能进行粗略地评价和改进。例如,检查表中可能出现这样的问题:“是否有将生产废料转化成有用产品的方法?”一个简单的回答“是”几乎没有携带任何信息。回收性能的改进还需要更进一步对废料的类型、回收的难易程度等进行分析 and 评价。
- 检查表可以检查的项目是有限的,检查表并不能使设计者关注那些重要的产品属性,例如,它无法指出减少资源用量和提高可回收性两者之间哪个更重要。也可能没有考虑产品研发经费或许会限制产品优化方案的实施。
- 仅仅通过检查表的一系列检查,设计者或许会误以为产品的设计已经符合要求而且考虑全面了。这样显然是不对的。所以说检查表只能作为设计初期对设计进行粗略分析的一种定性方法。

### (2) 矩阵分析 (Matrix Analysis)

矩阵法是一种对设计中多重决策进行折衷分析的有效方法。它通过矩阵的形式,用矩阵的行元素表示优化的目标,用矩阵的列元素表示产品的各种属性,对交叉项进行打分。矩阵法的最广泛应用之一就是质量屋。它通过将产品设计参数和用户要求进行综合评价、打分,从而优化参数和决策。在应用的过程中可对不同的优化目标进行加权,从而突出不同设计目标的重要程度。而且它还可以用于对产品生命周期的各个阶段进行分解、分析和综合。图 37.2-5 是矩阵法应用的一个例子。

### (3) 产品的生命周期评价 (Life-cycle Assessment)

生命周期评价是对产品在其整个生命周期内对环境所造成的影响进行评价的一种方法。使用这种方法

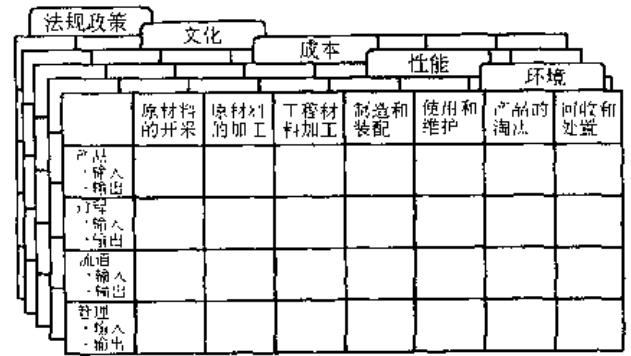


图 37.2-5 产品生命周期设计的分析矩阵

需要识别与量化产品在生命周期内消耗的能量与材料、释放的废物,分析它们对环境的影响并对改进方案进行评价。具体的过程包括:

- 目标定义与边界限定;
- 清单分析;
- 影响分析;
- 改进分析。

由图 37.2-6 可见, LCA 涉及了产品生命周期的各个阶段,对产品整个生命周期中向环境输入、输出的能量、材料与废弃物进行统计、折算、分析与评价。生命周期评价方法涉及范围的广度与评价内容的深度既是它的有力之处,也成为它的最大弱点。一方面,获得所需要的全部信息与数据是很困难的;另一方面,这些信息与数据的准确性、考察的目标、范围(边界)将对评价结果的有效性产生重要影响。目前,一些生命周期评价的软件工具已经出现,比如 SimaPro、IBIS、PRICE SEER 等。这些工具大都集中于清单分析,对各种层次和类型的能量与材料进行量化,而很少涉及影响分析和对改进方案的评价。

### ■ 影响分析 (Impact Analysis)

影响分析方法通过对产品在制造、运输、使用和处理过程中对环境释放的各种物质和能量进行分析,从而评价产品对环境造成的各种影响。这些影响包括对环境、健康、安全等方面的影响。这些影响可以是地方性的、区域性的,也可以是全球性的。例如对土地的使用和污染、对臭氧层的破坏、对使用者健康和安全的影响等。

由于知识的局限性,对一些环境恶化的规律性和成因还不甚明了,对环境变化和对健康危害的研究都还比较粗浅,因此对影响分析的应用和分析结果的可靠性都有待加深。但目前它仍然是绿色产品设计中经常采用的一种分析方法。

### ■ 环境会计 (Environment Accounting)

目前的会计方法并没有计入因环境管理的改善而引起的收入与支出的变化。这两个方面的原因,一是

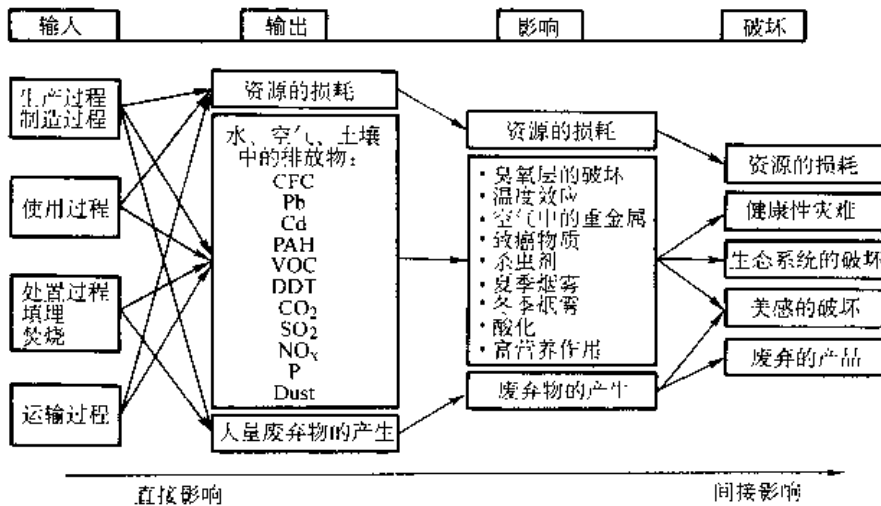


图 37.2.6 生命周期评价

传统的会计方法没有包括环境成本；二是环境成本的计算还很不成熟。例如，由于对材料、水、土壤和能量等资源的使用而造成对环境的影响是难以用经济指标来计算的。因为传统的价格计算是基于市场变化的，而市场中的供需关系是不能反映资源的使用和它的社会影响之间的关系的。也就是说通过提高产品的可回收性而带来的经济收益是传统会计方法难以计算的，即便通过当前回收材料的市场价格来计算，那也是不全面的，因为它没有计入那些间接的对环境的影响、对社会的影响，和随之带来的收益。

目前，基于活动的成本（ABC: Activity-based costing）、总成本评价（Total-cost assessment）和生态平衡（Eco-balance）等会计方法在面向环境的设计中得到了越来越多的应用。

### 2.4.3 设计方法

#### (1) 产品的生命周期设计 (Life-cycle Design: LCD)

产品的生命周期设计覆盖了产品的整个生命周期（见图 37.2-7），从需求分析、概念设计，直到使用、废弃、回收和处置。通常它采用生命周期评价、影响分析等方法对设计进行检查、分析和改进。在进行 LCD 设计的过程中要对产品生命周期的各个阶段进行分解和分析，对每一个阶段都要考虑以下几个方面：

- 成本
- 政策法规
- 保护环境
- 资源优化
- 产品性能
- 便于制造

也就是说，首先要明确产品在以上几个方面所应

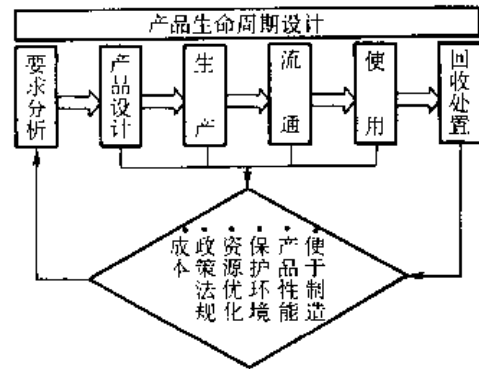


图 37.2-7 产品生命周期设计

该满足的功能、属性、质量等要求，然后再根据这些要求进行产品的概念设计、功能设计、详细设计、试制、批量生产、销售等。在设计、试制、生产乃至以后的过程中不断地发现问题并改进设计。

通常，LCD 采用生命周期评价或影响分析对产品的设计进行评估并做出修改。因此在设计的过程中要对产品生命周期各个阶段的输入输出（如材料、能量、劳动、废物等）进行分析。对于 LCD 这种方法来讲，“对产品要求的分析（requirement analysis）”和“产品生命周期评价”是设计成功与否的关键，也是 LCD 方法的核心。在 LCD 具体实施过程中，往往要使用多重矩阵进行问题的分解与分析。如图 37.2-5 所示，是一个简化的对产品进行分析的一个多重矩阵。事实上，对产品生命周期各阶段、各环节的分解可以根据具体情况而进一步细化。

#### (2) 面向 X 的设计 (DFX)

面向 X 的设计集中于对产品生命周期中某一具体阶段中某些具体特性的分析与改进，并常常以经济性作为评价指标。该类方法通常将产品的结构信息输入图表，计算各种方案与策略的成本与收益，进行经济

性分析与评价,从而得到最优的设计方案。目前的 DFX 方法一般只触及各类环境特性的某一方面,故缺乏整体性、综合性和系统性。知道了面向环境设计是基于产品生命周期的概念,因而可以按照生命周期各个阶段进行逐个研究,生成其 DFX 方法,然后再将这些 DFX 方法集成,最后形成综合考虑产品全生命周期的设计方法。具体地说,这些 DFX 方法包括面向制造的设计 (DFM)、面向装配的设计 (DFA)、面向制造与装配的设计 (DFMA)、面向维修的设计 (DFMAIN)、面向服务的设计 (DFS)、面向拆卸的设计 (DFD)、面向回收的设计 (DFR)、面向处置的设计 (DFEOL/EOd: End of Life) 等。

### (3) 生命周期设计与 DFX 方法的比较

前面已经讨论过有关面向环境设计的两类方法:基于生命周期评价的生命周期设计方法和各类 DFX 方法的集成。生命周期设计因其过程复杂、涉及的信息量巨大、信息的来源与评价未必能真实反应环境受到

的影响等原因,故而不适合一般的设计者在日常的产品设计中方便地使用。为此,生命周期设计与评价往往作为专门机构对产品的环境性能做总体评价,和对产品进行环境标志评定等状况下使用;有时也可用于大公司大集团对产品进行改进设计的时候。

从应用的角度来讲,基于各种 DFX 方法集成的绿色产品设计要更加实用。这些 DFX 方法也是基于生命周期各个阶段的,如上所述的面向使用和维护的设计、面向制造的设计、面向装配的设计、面向拆卸的设计和面向回收的设计等。它也可以是针对具体评价指标的,例如面向质量的设计、面向成本的设计、面向测试的设计等。表 37.2-1 列出了 LCD 和 DFX 两类方法的主要特点,两者在原理和基本的体系结构上有着本质的区别。当然它们也有相似与融合之处,比如都要提供生命周期中材料、过程、行为等输入、输出的数据清单;DFX 方法中有对全生命周期的考虑,而 LCD 方法中也有对可回收性、可拆卸性的要求。

表 37.2-1 LCD 与 DFX 方法比较

	生命周期设计 (LCD)	面向 X 的设计 (DFX)
基本思想方法	产品生命周期中各阶段能量与材料的输入输出对环境的影响	产品不同的功能、结构、材料、加工工艺、装配工艺、拆卸工艺及回收处置方法对环境的影响
评价方法	<ul style="list-style-type: none"> <li>产品生命周期评价,产品生命周期影响评价以能量为单位</li> </ul>	<ul style="list-style-type: none"> <li>单一评价产品的可加工性、可装配性、可拆卸性、可回收性、可维护性、可靠性与耐用性等</li> <li>以经济性指标(成本/利润)为单位</li> </ul>
数据库	<ul style="list-style-type: none"> <li>材料、能量、废弃物</li> <li>过程(制造、使用、处置)</li> <li>材料、过程消耗能量的折算</li> <li>材料、过程对环境影响的折算</li> <li>数据、资料来源、可靠性</li> <li>标准、法律、税收制度等</li> </ul>	<ul style="list-style-type: none"> <li>材料</li> <li>产品结构、联接</li> <li>过程(制造、使用、处置)</li> <li>过程消耗的时间,成本-利润</li> <li>标准、法律、税收、市场等</li> </ul>
应用对象	<ul style="list-style-type: none"> <li>复杂的大型 LCD 工具往往适用于专业环保人员,产品环境性能分析人员,大型公司、企业、研究机构</li> <li>产品设计人员常采用简化的 LCD 或指南、手册等</li> </ul>	<ul style="list-style-type: none"> <li>企业、公司、研究机构</li> <li>设计人员</li> </ul>
对产品设计的优劣势	<ul style="list-style-type: none"> <li>系统、全面,但对设计人员环境专业知识要求高</li> <li>较为抽象(采用能量单位,评价环境因素如臭氧层)</li> <li>难以对结构设计提出改进</li> </ul>	<ul style="list-style-type: none"> <li>具体、有针对性</li> <li>直观</li> <li>便于发现具体问题</li> </ul>

通过以上的介绍、分析和讨论,可以看出,绿色产品设计是近年来发展起来的一种基于产品全生命周期的系统的设计方法。从基本的原理和方法上可以把现有的绿色产品设计方法分成两大类,一类是生命周期设计,另一类是分布于产品生命周期各个阶段上的 DFX 方法的集成。这两类方法各有特色,生命周期设计是对产品在生命周期各个阶段上对环境的输入输出、造成的影响进行综合评价,进而优化产品设计,以保证设计出来的产品对环境的综合影响最小。然而,这种方法的实施需要大量的数据,和对环境、污染、资源等方面的较为全面的知识,这对于一般的设计人员来讲是有一定难度的。因而生命周期设计较适用于一些大型的公

司,和对较为成熟的产品不断更新改进设计上。而 DFX 方法因其较为具体,对数据和设计人员的要求就没有生命周期设计方法那么严格,这也就使得它更加实用。但从另一方面来讲,DFX 方法的系统性和综合性就不如生命周期设计方法,因而对于 DFX 方法的集成就显得更加重要了。这就要求它们具有统一的框架体系和信息系统的支持。绿色产品设计虽然近年来得到了一定的发展,但还处在其发展时期。要使它成为一种系统的广泛应用的设计方法,似乎还有较长的一段路要走。人们应当看到,21 世纪面临着环境、资源、人口等危及人类生存和发展的诸多挑战,要实现可持续发展,必然要从以往环境保护的末端治理走向源头预

防。对制造业来讲,要根本地解决环境问题就要依靠绿色产品设计。

## 2.5 绿色产品设计的工具

从上一节所分析的传统产品设计和绿色产品设计的区别中,可以看出,传统产品设计人员无需了解所设计的产品与环境之间的基本关系;然而一个绿色产品设计人员,则必须懂得他所设计的产品与环境之间的基本关系。为了使产品设计人员能用一种充分考虑环境负担的方式去开发产品,他就必须学会从环境负担的角度去评估:哪种产品设计方案更好。为此,需要为绿色产品设计人员提供相应的环境负担的评估工具。

客观上,有着各种可能的评估方法达到这一目的。此处主要着重介绍:

■ 采用类似设计准则的定性方法

■ 采用生态指标(Eco-indicator)的简化定量方法

■ 采用生命周期评估(LCA: Life Cycle Assessment)的定量方法

上述这3种方法的实施难易程度和评估精准程度,是依次递增的。以下分别由易及难、由粗及精地介绍这3种方法。

### 2.5.1 供设计人员评估其产品环境影响用的定性准则

一般地说,一个产品设计人员无需作深入的分析,便可评估出所设计产品的环境影响。然而,这时他需要掌握一些基本的评估环境影响的定性准则。下面便给出10条基本准则。这10条准则可以用来评估大多数产品的环境影响。这是因为这些准则基本上体现了绿色产品设计的生命周期思想。

(1) 不是单纯按功能设计产品,而是要强调从环境上按产品生命周期来设计产品

这一准则要求设计人员在设计产品时,应考虑产品生命周期的全过程,也即从构思设计产品开始,到制成产品后经使用、维修而最终报废为止。这在国外被称作“从摇篮到坟墓”,即所谓的Cradle-to-grave;或者经使用、维修而报废后被回收、再生、重用,这在国外称作“从摇篮到摇篮”,即所谓的Cradle-to-cradle。绿色产品设计强调的是后者。

为了考虑产品全生命周期的环境影响,此处提出一种如表37.2-2所示的“材料-能量-毒性”矩阵表,又称MET矩阵表(MET-Material, Energy, Toxicity)。它是记录设计人员设计结果的一种简单方法。这时只需将最重要的事实填写在下面所示的矩阵表中即可。

(2) 并非天然材料总是好的

一般总认为“天然”材料要优于“人造”材料,其实不然。虽然生产1kg木材的排放物要比生产1kg塑料少得多,但是有否考虑过:木材需从原木经过加工,然后才能得到所要的木材制成品,而木材制成品还要经过油漆?这时必须考虑锯木厂、木工厂、油漆厂各自的MET矩阵表。众所周知,塑料是可以回收的,可是木材则较难回收。这一准则要求设计人员能对所选用的材料在1kg的基础上作比较。因此,必须收集各种材料1kg时所消耗的能源数量、毒性程度、各种排放物的数量等数据,并编订成手册,以供设计人员在评估环境影响时使用。

表 37.2-2 MET 矩阵表

生产周期阶段	材料	能量	毒性
生产			
使用			
报废后处置			

(3) 不要低估能源的消耗

许多设计人员在设计产品时,把注意力集中在材料选择上,而恰恰忽视了所设计产品的能源消耗。如果一种产品在其使用过程中要消耗能源,那就有能源消耗占优势地位的可能。通常人们总是低估了能源的环境影响。

下面列举的几个简单例子中的数据,或许会对理解能源的环境影响有所帮助。

- 每10kw-hr的电要用2kg石油;
- 每1kg塑料要用1.5~2.5kg石油;
- 煮咖啡机在其寿命期内,要用掉300kW·h电力,相当于60kg的石油。然而在生产煮咖啡机时,只用了2kg塑料,相当于3~5kg的石油。

从煮咖啡机的实例中,可以深刻领悟能源的环境影响;制造煮咖啡机只消耗3~5kg石油,而使用煮咖啡机直至其寿命终止时,却消耗掉60kg的石油。在这一例子中,产品使用阶段的能源消耗,明显占有优势地位。由此可见,不能只考虑产品设计阶段的能源环境影响,而必须考虑产品整个生命周期的能源环境影响。为此,产品设计人员切莫低估能源消耗。

(4) 提高产品的寿命

产品设计人员可以通过不同途径去提高产品寿命。例如:

- 从技术上使产品更耐用(采用新材料和新工艺等);
- 从结构上使产品能更新(采用标准接口、模块结构等)。

必须指出的是,今天有不少产品被人们抛弃,并非因为它已损坏,而恰恰只是因为它的功能落伍、外观令人生厌。因此,重要的是设计人员能将产品设计成能深深地吸引住人,令人感到缺它不可或非它莫属。这就是



从人们的心理上去延长产品的寿命。比如现在的手机(移动电话),由于它的技术进步太快,因而导致它迅速更新换代。老手机被淘汰的原因,往往不是它的物理寿命已经终止,而只是它的技术寿命已终止。

(5) 不是单纯设计产品,而是通过设计产品能给人们提供更广泛的服务

人们并非总是想要通过买断产权(产品所有权)而独占一个产品。因为有些产品并不经常使用,所以他们只需要临时租用产品的使用权即可。正因为这样,人们有时候真正需要的,是对某个具体问题的一种解决办法。因此,“提供广泛服务”的这一观念,可能是一种好的解决办法。

例如:建立公共洗衣机房,就是解决洗衣问题的一个很好的办法。这样,建立一个公共洗衣机房,其实就可以替代许多台个人洗衣机。它不仅节约了制造这些被替代的资源,而且还充分体现了“物尽其用”的另一种节约资源的理念。因为若不设立公共洗衣机房,则每家每户势必都要购置单独的洗衣机,但是这些单独的洗衣机并不能充分利用,大部分时间是被闲置不用,而这种“物不能尽其用”,也是另一种浪费资源的表现。

类似上述单独洗衣机的使用情况。轿车也是一种消耗资源较多而又未能充分利用的产品。据报道,为了改善轿车的利用情况和城市的交通状况,荷兰在试行各种轿车共享制度。共享制度的核心,就是通过合理的计划安排,使轿车按时段的原则,分配给同时段的人们共用;于是,轿车既被充分利用,而人们又能分时共享。这样便可控制车辆数量,提高车辆的利用率,减少停车的车库和泊位所占用的土地,减少车辆的尾气排放,减少能源的消耗,大大改善城市交通拥挤堵塞的状况。

综上所述,有时换一种思路考虑产品设计,不从单纯的设计角度,而是从产品的使用角度,考虑如何使产品更好地服务,更充分地被利用,也能达到节约资源、减少污染的目的。

#### (6) 用料最少

这条准则似乎是显而易见的,因为绿色产品设计就是要求制造产品所用的材料为最少;但是真的要作到这一点,却要比想象的复杂得多。一般通过严格的尺寸和强度计算,通过相应的生产技术,是可能降低材料的总量的。如果能够使材料总量减少,即使采用每1kg具有高环境负荷的材料也是可考虑的。这一点对于设计运输设备之类的产品尤为重要。运输设备的自重愈轻,则它所消耗的燃料也愈少。生态指标法适合于作这种评估。

#### (7) 尽可能使用回收再生材料

产品设计人员不仅使自己所设计的产品是可回收

的,而且还应该在产品设计中积极采用回收后再生的材料。如果每个设计人员只考虑产品是可回收的,而不考虑在设计中采用回收后再生的材料,那么产品回收后回炉再生的材料就永远不会有需求。显然,只有当对回收后再生的材料有需求,然后才肯定会有人组织供应。

当然,在产品设计中并不限于只采用回收后再生的材料,也提倡采用回收后再制造的零件和部件。这既能大幅度地降低成本,同时又能显著节约资源。因为再制造并非从毛坯或原材料开始重新加工,而是在回收后原有的零部件基础上加以修复翻新,所以采用再制造的零部件能够降低成本和节约资源。

#### (8) 使所设计的产品是可回收的

应该说,大多数产品是可以回收的,但是实际上目前却只有少数产品被回收。其原因就在于选作被回收的产品必须满足下述要求:

- 产品本身属于量大面广的;
- 产品易被拆卸的。

这里主要考虑的是回收的经济性。只有当产品的产量很大,回收的意义和价值才愈大;只有当产品容易拆卸,拆卸的成本才能最小。产品设计人员可以通过优化产品的结构,使之易于拆卸而增加产品可回收的机会。

#### (9) 从废旧产品中回收材料

这一准则体现了绿色产品设计的基本思想之一,即在废旧产品中的零部件无法回收再制造时,则最终只能将其作为再生材料回收回来。这样便能最大限度地重复利用资源而不致浪费。

从废旧产品中回收塑料、钢料和织物等时,应记住下述准则:

- 如果想要回收热塑性塑料,则必须注意
  - 不要在热塑性塑料表面上涂刷油漆;
  - 不要在热塑性塑料表面上粘贴粘胶标签;
  - 不要混有其他塑料。

因为热塑性塑料表面上所刷的油漆或所粘贴的粘胶标签等,是极难清除干净的;所以若一旦混入热塑性塑料后,则将影响所回收再生材料的品质和色泽。

- 若要回收钢料,则须注意在熔化时不可混入过多的铜。

- 对于热固性塑料和织物,因回收后并无用处,故建议回收后焚烧,如此尚可从中收回热能。

(10) 设计人员在设计产品过程中多问自己几个“为什么?”

许多情况下,决策往往是基于常识作出的。人们的确经常是这样作出决策,而且一般也很奏效。不少设计人员就是通过不断反复向自己提问一些简单的,甚至是愚蠢的“为什么?”的问题,而获得有效改进产品环

境影响的答案。这里不妨举例说明这条准则的作用和效果。

- 一家包装公司设计制造一种包装。按要求,应保证所包装的内容物必须保鲜 18 个月。设计人员困惑不解的是,这种内容物“为什么必须保鲜这么长时间?”。于是设计人员决定进行细致深入的调查研究。最后发现,所包装的内容物一般不超过 3 个月就消费掉了。因此,设计人员便采用完全不同的另一种包装设计。这样不仅简化了包装,降低了包装成本,并且还节约了资源。
- 有家公司用了 3kg 原料制得 1kg 聚合物。后来研制人员反复提问“为什么用料这么多?能否减少投入,增加产出?”。经过试验,终于找到另一种方法,可以使原料浪费少于 0.5kg。
- 据报道,荷兰的雨水管道(落水管)直径,传统上都是采用  $\phi 80\text{mm}$ 。有人提出“为什么要用如此粗的管径?有此必要吗?”。后经研究荷兰的房屋建筑条例,并对雨水入口的流体力学性质作了简单优化,发现落水管管径采用  $\phi 30\text{mm}$  就已足够。由于显著缩小房屋建筑所用的落水管管径,因此仅此一项,就可节约制造落水管所需的大量资源。

### 2.5.2 生态指标法

设计人员并不是一个环境专家,而且他们也永远无必要成为一个环境专家。然而,设计人员在设计产品和选择制造方法时所作出的决策,却严重地影响着产品全生命周期期间的环境。显然,设计人员也不可能对每个环境问题,都去咨询环境专家。因此,有必要为设计人员提供一种实用可靠的工具,以评定他们所作设计决策的环境后果。正是由于这种考虑,国外已经开发出多种此类工具。这里我们从中选择一种较为成熟的、由荷兰一家专门从事绿色产品设计和环境影响评估软件研发的 Pré 咨询公司所开发的工具:生态指标(Eco-Indicator),作一介绍。

生态指标是一种用来计算常用材料和加工方法的标准指标分值的方法。这些标准指标分值,可以从生态指标 95 设计手册中得到。这本手册中最初的 100 种常用材料和加工方法的生态指标分值,是在生命周期评估方法(LCA: Life Cycle Assessment)的基础上,用 Pré 咨询公司会同飞利浦(Philips)等公司、德尔弗特(Delft)等大学以及荷兰金属工艺研究所(TNO)等咨询机构合作开发的。这一项目也是荷兰政府下属城市、住房与环境部所资助的。

产品设计人员有了生态指标这样的工具,就可以

对自己所设计的产品,进行环境影响方面的评估。由此就能从不同的产品设计方案中,选出环境影响最小的绿色产品。

#### (1) 用生态指标作绿色产品设计

每种产品都会在其生命周期的某个阶段,对环境带来各种影响。例如,制造产品必然要提炼原材料,必须包装和配送产品,最后还要处置被废弃的产品。特别是在产品使用阶段,它要消耗能源,而且产品中的零部件也会有耗损。由此可见,为了评估一个产品的环境影响,就必须对产品生命周期的所有阶段进行分析和研究。对产品全生命周期环境影响的分析研究,便称为产品的生命周期评估。目前,生命周期评估主要有两种用途:

- 用于产品设计时的方案比较和选用。这时主要为了确定各个方案的总的环境影响,从而选出总的环境影响最小的人选方案。当然,这并不限于产品的总体设计方案,实际上也可用于零部件备选方案或材料备选方案的比较与选用上。
- 用于确定造成产品给环境带来负面影响的最重要的原因。由此,设计人员即可集中精力消除这些原因,从而达到改善产品的环境性能,提高产品的绿色度。

虽然产品设计人员希望将生命周期评估用于产品的设计过程,但是如今他们却面对着两个主要问题:

- 难于解释生命周期评估所得出的结果。通过生命周期评估,尽管可以确定暖室效应、酸化、重金属以及其他等因素,对产品生命周期的环境影响所起的作用;然而,对总的环境影响却仍为未知。究其原因,实乃缺乏各影响因素间的相互权重。
- 一般而言,仔细收集一个产品生命周期的全部环境数据,是既复杂而又费钱、费时的;因此,通常在产品设计过程中,是无法实现扩大的生命周期评估的。

好在生态指标法已经解决了这两个问题:

- Pré 咨询公司已将生命周期评估方法加以扩充,使之包括一个加权方法。这就有可能基于计算所得的各个影响因素的结果,对总的环境影响去计算一个单一的分值。这个分值的数字,便被称作:生态指标。
- Pré 咨询公司已经收集了大多数常用材料和工艺方法的环境数据。生态指标正是由这些数据计算面得。所有的材料和工艺方法,就如砌墙的砖那样,彼此搭配默契。因此,就有生产 1kg 聚乙烯的(材料)生态指标,挤出 1kg 聚乙烯的(工艺方法)生态指标,乃至焚烧 1kg 热塑

性塑料的(最终处置)生态指标。

一种材料或一种工艺方法的生态指标,其实只是在生命周期评估的数据基础上所得到的一个数字,而它却指明了该种材料或该种工艺方法的环境影响。显然,生态指标愈高,则其对应的环境影响也愈大。由此可见,生态指标把对环境的评估带入到产品设计人员的设计过程中。

### (2) 生态指标工具的应用

有了生态指标后,生态指标的应用是比较简单的。首先,设计人员必须列出所设计产品在其生命周期内

出现的各种材料、能源和工艺方法的总需要量。即如前面所提到的,要编制一份MET矩阵表。然后查找这些材料和工艺方法的生态指标分值,并且将各自的总需要量分别乘以相应的生态指标分值。在完成上述计算后,即可分析出何种工艺方法对整个环境影响所起作用最大。接下来,便是寻找各种变通的设计方案,同时从环境的观点去分析这些方案是否真的更为可取?下面举一个应用实例,介绍Pre公司就煮咖啡机这一产品,利用生态指标所作的环境影响分析的结果,如图37.2-8所示。

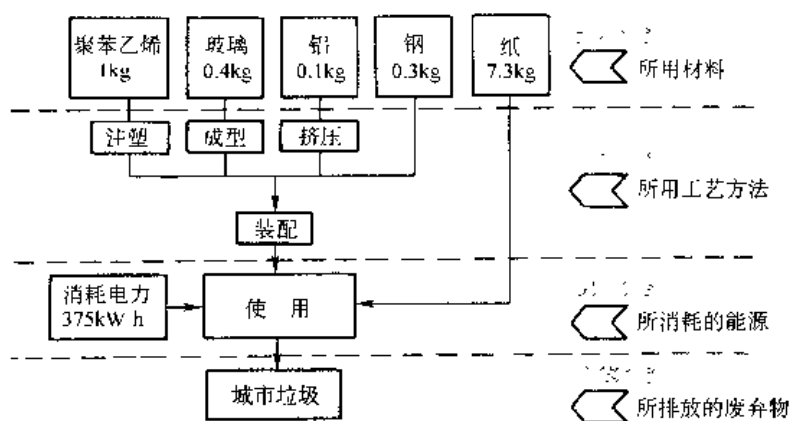


图 37.2-8 煮咖啡机全生命周期环境影响分析的数据与资料

图中的每个方框,都可以从生态指标手册上,查到其相应的生态指标分值。根据所查出的生态指标分值,可以看出,在煮咖啡机的全生命周期中,纸质过滤袋和加热所消耗的电力,是对环境影响起决定作用的两个主要因素。因此,设计人员必须考虑设法减少纸和电的消耗。一种可供选择的方案,是采用保暖壶的方式,这样,便能取消为使咖啡保暖而消耗的电力。于是,按此方案,设计人员即可计算出生产保暖壶的环境负担,并分析如何借此去平衡所减少的电力消耗。

### (3) 生态指标的计算

生态指标的关键点,是其计算的透明度。这是因为人们需要知道指标的意义;同时在必要时,也能够去计算新的指标。有关生态指标的论述,可从Pré咨询公司的网页下载一份相应的报告中了解;有关生态指标的计算方法,则可从Pré咨询公司开发的软件Simapro中获得。

在开发生态指标时,Pré咨询公司的研究人员尽可能采用与生命周期评估方法学相接近的方法。有关这种方法的叙述,可见SETAC(Society of Environmental Toxicology and Chemistry,这是总部设在比利时布鲁塞尔的一个学术团体:环境毒物学与化学学会)的出版物。之所以要尽可能采用与生命周期评估方法相近的方法,是因为希望采用生态指标方法进行分析的结果,应该与生命周期评估方法所获结果

尽可能相同。在此强调这一点,是想说明生态指标方法的初始阶段,是与生命周期评估方法的步骤相同的。

#### ● 编制环境负担因素清单 (Inventory)

正如以上所述,在Pre公司的Eco-indicator 95版本中,最初仅仅只考虑了100种常用的材料和工艺方法。就这些环境负担因素构成的清单进行了生命周期评估。而后又对另外近100种材料和工艺方法进行了生命周期评估。于是通过修订,而有了Eco-indicator 99这第2个版本。这样,几乎把所有有关环境负担因素都分析到了;同时也收集了所有的排放物,以形成一份环境影响表,因而能对排放物有一个总的概貌。

#### ● 对环境影响后果分类

在上述清单分析所得环境影响表的基础上,对大量环境影响后果进行了计算。通过对环境影响后果的分类,就有可能对两个同类产品的环境影响后果进行比较,如图37.2-9所示,即为上述这种比较的常用表示方法。在此,将环境影响后果划分为如图中所示的8类。

图37.2-9也反映了煮咖啡机中所用聚乙烯过滤袋和纸质过滤袋二者各自生命周期内环境影响的特点。

至此,生态指标是完全遵守经典的生命周期评估方法。然而,在此实例中,很难对其结果给出解释。从图37.2-9中可以看出,纸质过滤袋除了产生较多的致癌物质外,它的其他环境影响后果的分值还是较好的。

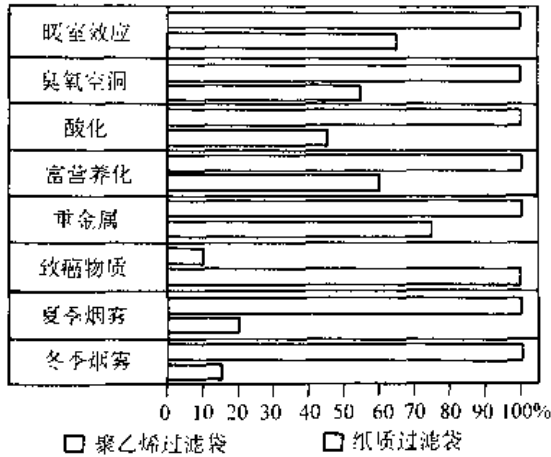


图 37.2-9 在聚乙烯过滤袋与纸质过滤袋之间的比较实例 (此处设定每种环境影响后果的最高分值为 100%)

由此可知,生命周期评估并未揭示出哪种过滤袋更好。这里所欠缺的,是上述环境影响后果之间的权重。虽然生命周期评估方法阐述了关于应如何加权的问题,但是由于缺乏必要的数据,因此也几乎没有实现加权的步骤。正是生态指标的研究项目,填补了这一遗漏。

(4) 规范化和评价

在图 37.2-9 的基础上,不大可能确定哪一种过滤袋对环境的影响更小。首先,因为两者所得的高分值均以 100% 为尺度。显然,实际上这种尺度是无意义的。因为分值为 100%, 可以代表一个很小的或很大的排放量;所以在作任何更深入解释之前的第一步,必须将图 37.2-9 中的这些分值与另外的分值比较。在生命周期评估中,这被称作规范化。Pre 咨询公司在开展生态指标项目的研究中,为了实现这一规范化而开发了所谓的“居民当量”。即平均一个欧洲人在一年内所造成的环境影响后果。由此,先将第一步的分值规范化为一个欧洲居民当量,如图 37.2-10 所示。如今,环境影响的后果便可在居民当量的尺度上进行比较。正因为如此,很明显,从绝对意义上看,“臭氧层空洞、富营养化、杀虫剂和致癌物质等的分值就很低;而两类烟雾的分值和酸化、重金属以及暖室效应的分值便相对较高。

规范化从相对的意义,揭示了何种环境影响的后果大,何种环境影响的后果小;但是它仍然未说清楚各种环境影响后果的相对重要性。一种很小的后果,也许是最重要的。因此,为了得到一个总的结果,就需要一种加权的步骤。加权的步骤即如图 37.2-11 所示。关于用于这最后评价步骤的加权因子,将在下面进行讨论。

至此,所有的环境影响后果都用一定的严重性尺度标定。在这个例子中,严重性使用生态指标的点数 ( $\mu Pt$ ) 来表示。如果在图 37.2-11 中,沿水平方向按生

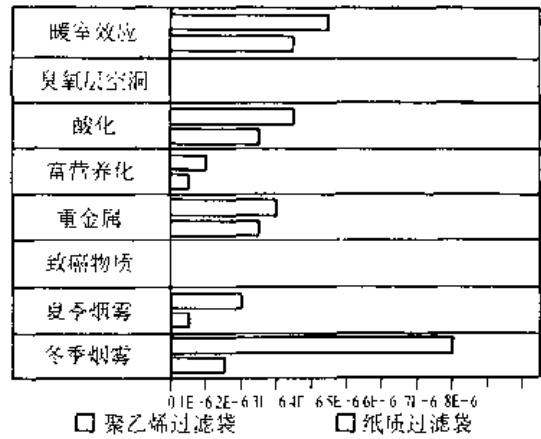


图 37.2-10 此处将来自图 37.2-5 的环境影响结果分值,规范化为一个欧洲居民当量 (即该欧洲居民 年内 1000 个过滤袋造成的暖室效应后果)

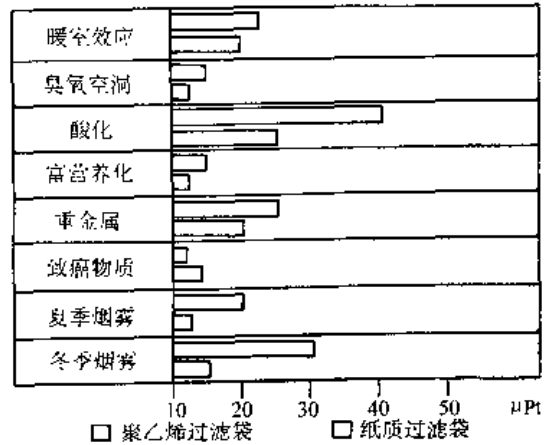
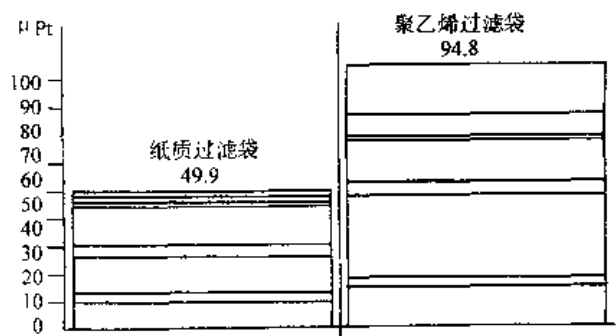


图 37.2-11 评价步骤:对规范化后的环境影响后果的分值加权

态指标  $\mu Pt$  相同尺度画出所有的直方条,则这些直方条的长度 (即生态指标点数),原则上,就可累加求和。累加求和的结果,便是如图 37.2-12 所示。在图中纸质过滤袋的生态指标的点数为 49.5,聚乙烯过滤袋的生



(上图中的直方条自下而上分别表示:暖室效应、酸化、富营养化、重金属、致癌物质、夏季烟雾和冬季烟雾。)

图 37.2-12 累加求和结果

态指标点数为 94.8。由此可以清楚地看出,纸质过虑袋在对环境友好性上,要比聚乙烯过虑袋略差一筹,尽管两者的差别不是很大。

#### (5) 加权的背景

根据以上这些图,似乎环境影响后果的加权是非常直截了当的。其实,问题还在于加权因子是如何确定的。在生态指标的研究项目中,对于这个问题考虑得很多。经过对各种可选方法的详细分析,最后选择了所谓“距离到目标”的方法来确定加权因子。这一原理在瑞士的生态计分(Swiss Ecopoints)加权系统中已经用了多年。这个方法的基本前提是,在当前水平和目标水平间的距离与一环境影响后果的严重性之间,有着一种相关性。因此,为了达到一个可持续发展的社会,必须使酸化因子降低 10,而烟雾因子则必须降低 5。这样,酸化的严重性便看成是烟雾的两倍。这里的降低因子也就是加权因子。在 Pre 咨询公司的生态指标项目中,他们对这一原理作了推敲和改进。此处限于篇幅,无法就此展开详细阐述。有兴趣的读者,可阅读本篇所附的参考文献。

“目标水平”这一术语仍然是一个问题。什么是一个好的目标水平?这样一个水平又是如何确定的?上面提到的瑞士生态计分方法,采用的是来自政府政策报告中的政治目标水平。这类水平通常是根据可行性(成本)与愿望之间的一种折中来确定的。在 Pre 咨询公司生态指标研究项目中,研究人员是这样来定义目标水平的,即目标水平独立于政治之外,而且是以科学信息为基础。然而问题并未真正解决,因为科研人员对于建立怎样一个良好的目标水平,有着不同的观点。不同的环境问题引发不同类型的环境损害。例如,对于烟雾而言,引起健康方面的抱怨;而对于酸化,则主要造成森林的损害。为了确保对酸化的目标水平相当于对烟雾的目标水平,故而必须按环境影响后果所引起的损害,建立一种相关关系。建立这种相关关系的前提是,对于

每一种环境影响后果的目标水平,都应一律生成严重的损害。下面所列出的损害水平假设是等量的:

- 把灾害的数量作为环境影响后果的一种重要结果。因此,把每年每百万居民的一个灾害选作可接受的水平。
- 把生病的病人数作为环境影响后果的一种重要结果。这尤其是对夏季和冬季的烟雾具有重要意义。设定的可接受水平为,烟雾周期应很难再次出现。
- 生态系统退化。已经选定的目标水平为,在此水平上,要经过几十年才能继续出现 5% 的生态系统退化。

为这些损害的水平设定当量,是一种主观的选择。它们不可能作为有科学意义的根据。因此,就有可能作出不同的假设,这样也就导致加权因子有变化。目前这类当量的选择,是来自不同专家的咨询以及和其他系统的比较,如瑞典的 EPS 系统等。图 37-2-13 是生态指标原理的一个示意性的框架描述。

为了建立这些损害水平与影响后果之间的一种相关关系,Pre 咨询公司进行了一项有关欧洲环境实际状况的详细研究。于是,首先要确定每种影响后果的现状;同时,对于一具体的影响后果而言,必须降低到何种程度,方可达到为其确定的损害水平。为此 Pre 咨询公司,尤其是荷兰国家公共健康与环境卫生研究所,在此领域中做了大量工作。如今已经可以得到一张详细表示环境问题的欧洲地图。这些数据用来决定某一环境问题的当前水平,同时必须使该问题降低到某个因子,以达到一可接受的水平。表 37-2-3 列出了加权因子和所采用的准则。在这张表上,可以明显看出,应该优先加以限制的,是导致臭氧层损害的和采用杀虫剂的物质。尤其是杀虫剂,这在荷兰是一个非常严重的问题。再者,应该更多关注酸化和致癌物质的扩散。

- 有毒物质仅仅是一个出现在工作场所的问题

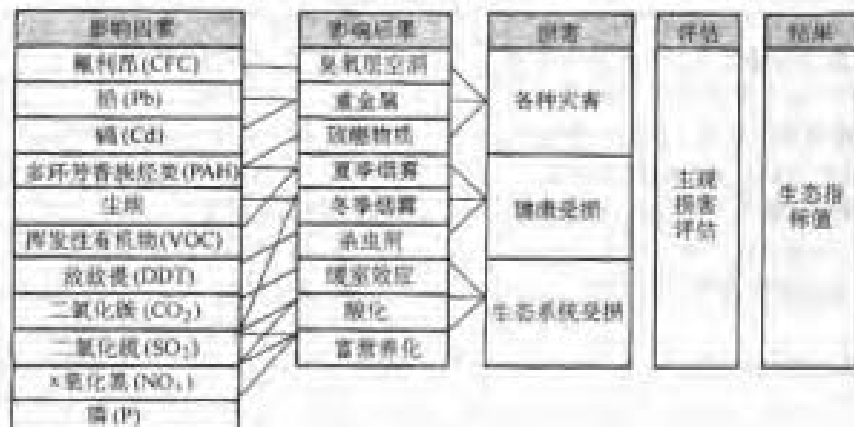


图 37-2-13 生态指标的加权原理

因为许多物质只有当它们出现某种集中时,才是有害的;而这种有害的集中,比较容易出现在工作场所。如果在这集中区以外,则往往毒性遗留较低,而且低于伤害的阈值。之所以会这样,是由于这类物质一般都被大大地稀释掉,同时许多物质通过自然分解过程而在大气中消失。只有那些在有害集中区找到的物质,则都包括在生态指标中;其余的,均被忽略不计。这就意味着,具有低生态指标分值的产物,仍然会由于释放局部有害的物质,而导致恶化工作条件。

表 37.2-3 环境影响后果的加权因子和所用准则

环境影响后果	权重	准则
暖室效应	2.5	每 10 年上升 0.1°C, 生态系统退化 5%
臭氧层空洞	100	每百万居民每年 1 种灾害的概率
酸化	10	生态系统退化 5%
富营养化	5	河流湖泊, 数量未知的水中生态系统的退化 (退化 5%)
夏季烟雾	2.5	烟雾出现周期, 健康抱怨 (特别是哮喘病人和老人), 防止农业损害
冬季烟雾	5	烟雾出现周期, 健康抱怨 (特别是哮喘病人和老人)
杀虫剂	2.5	生态系统退化 5%
空气中的重金属	5	儿童血液中的铅含量, 在数量未知的人群中降低了预期寿命和学习能力
水中的重金属	5	河流中的镉含量, 最终也将影响到人 (见上述空气中的重金属)
致癌物质	10	每百万居民每年 1 种灾害的概率

● 原材料的耗费

如果是由稀有原材料制成的一种产品, 则这种稀有性不会表现在指标中; 毕竟稀有的物质不会引起对健康的任何伤害。由于必须始终采用低品位的矿石, 所以指标中包含了提炼原材料时所产生的排放物。这种排放物通常还是量很大的。顺便指出, “耗费” 这一术语还很难定义。对于大多数原材料而言, 可有各种方案能使原材料保持更长时间的循环利用。事实上, 材料不会从地球上消失; 作最坏的打算, 它们也就是以一种不适宜的方式扩散掉。

● 废弃物

其实, 在环境关系中, 废弃物占据的空间并不特别重要, 因为堆积如山的废弃物使生态系统损失的量, 比起如酸化所引起的对生态系统的损害, 相对而言要小得多。然而, 废弃物所释放的物质 (重金属或焚烧时的 CO<sub>2</sub>) 却十分重要。这后者的环境影响后果均包含在指标中, 只是废弃物本身的数量并非评估过程的部分。

由于这些差异, 生态指标可以看成是一种排放物的指标, 至于原材料的耗费和废弃物的空间使用, 目前必须分开评估。

从以上所介绍的生态指标法中, 可以看出, 生态指标基本上是供产品设计人员, 在设计过程中所使用的一种工具。它使产品设计人员, 利用 Pré 咨询公司已对 200 种常用材料和工艺方法所制订的生态指标, 作出

自己的产品生命周期评估。产品设计人员可按下述两种方式运用生态指标:

● 得出正确的问题 构成一个产品环境负担的基本原因是哪些?

● 得到正确的答案 哪个方案的环境负担最低。

这一方法是 SETAC 的 LCA 方法的扩展。它采用了规范化步骤和评价步骤。这一评价的根据是, 最为可用的、一种欧洲尺度的环境影响后果损害的知识。

2.5.3 生命周期评估法 (LCA)

一个产品与环境之间的复杂互动关系, 是在产品的生命周期评估方法中处理的。下面便对这一方法作扼要介绍。LCA 有两个主要的步骤, 即:

● 描述在一个产品的生命周期中, 要用到哪些原材料, 以及将出现哪些排放物。这就是通常所谓的编制环境影响因素清单步骤。

● 评估这些排放物与原材料耗费的环境影响是什么。这就是所谓的环境影响评估步骤。

在产品设计中要采用 LCA 的原因, 就在于: 政府和顾客只希望各个公司应关注所有各种产品的环境品质。EMAS、BS 和 ISO14000 系列标准, 则要求各个公司不断改进自身的环境管理系统。而 LCA 及其应用, 却正好符合对产品/工艺方法改进的这一要求。这里介绍的 LCA 方法, 主要是根据莱顿大学 CML 和环境毒性学和化学学会 (SETAC) 所提出的。在 SETAC 的“实践规范”中, 它将 LCA 划分为 5 个阶段:

● 规划阶段

- 陈述目标
- 确定产品及其各个方案
- 选择系统的边界
- 选择各个环境参数
- 选择组合和评价方法
- 数据收集的方针

● 筛选阶段

- 初步实施 LCA
- 调整计划

● 数据收集和数据处理阶段

测定、面谈、文献搜索、理论计算、数据库搜索、有资质的测算

- 环境影响因素清单的计算

● 评价阶段

对环境影响因素表分类, 将其划分成影响类别

- 在类别内组合 (特性化)
- 规范化
- 不同类别的加权 (评价)

### ● 改进评估阶段

#### □ 敏感性分析

#### □ 改进优先性和可行性的评估

在以上5个阶段中,普遍认为第1阶段是最重要的。LCA的结果很大程度上取决于这一阶段所采取的决策。筛选是检验目标确定阶段的有用步骤。经过筛选之后,将更加容易去安排项目其余的计划。下面将就环境影响因素确定和环境影响评估,作更进一步的介绍。

#### (1) 产品生命周期的环境影响因素清单

LCA研究的基础,是一份出现在一个产品生命周期期间的、全部工业过程的输入与输出的清单。这份清单不仅包括产品的生产阶段,而且还包括产品的配送(包装与运输)、使用和最终处置等生命周期的全过程。产品的生命周期可以用如图37.2-14所示的一棵过程树来表达。图中的每一方块表示一个过程。每个过程构成产品生命周期的一个部分。每个过程也都有自己已经确定的输入和输出。

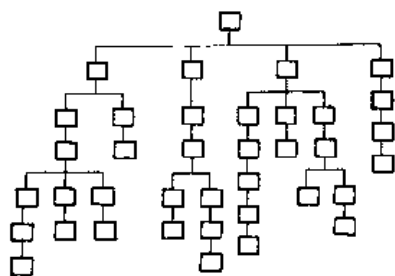


图 37.2-14 表示产品生命周期的过程树

过程的输入可以划分为两部分:

- 原材料与能源等资源的输入(也有人称之为:环境的输入);
- 产品、半成品或能源的输入。这些都是来自其他过程的输出(也可以称作:经济的输入)。

同样,输出也可以划分为两类:

- 排放物的输出(又称:环境的输出);
- 产品、半成品或能源的输出(又称:经济的输出)。

利用上述产品生命周期的过程树,和每个过程的有关信息,我们就能编制出与该产品全生命周期相联系的、所有环境的输入与输出的一份清单。这一清单的结果,便称作环境影响表。每一环境影响都被表示为一种物质的具体数量。作为示例,表37.2-4展示了一张环境影响表的一小部分。表中显示的是两种材料(玻璃与聚乙烯)的环境影响。对于这两种材料的环境影响表而言,若要完整地列出,则有34行。显而易见,单靠这样一张表是无法立刻就下述问题得出明确的答案的:生产1kg玻璃是否就比生产1kg聚乙烯的环境影响更小?正因为如此,所以就要开发环境影响的评估方法。它能简化回答上述问题的任务。在讨论环境影响的

评估方法以前,先要考虑和研究有关环境影响表的某些计算问题

#### (2) 编制环境影响因素清单过程的细化

表 37.2-4 生产1kg玻璃和1kg聚乙烯所产生的某些环境影响

排放物	玻璃	聚乙烯	单位
CO <sub>2</sub>	0.4901	1.792	kg
NO <sub>x</sub>	$1.586 \times 10^{-3}$	$1.091 \times 10^{-3}$	kg
SO <sub>2</sub>	$2.652 \times 10^{-3}$	$987.0 \times 10^{-6}$	kg
CO	$37.00 \times 10^{-3}$	$670.0 \times 10^{-6}$	kg

从原理上说,编制环境影响因素清单的过程,似乎很简单;然而,在实践上,却遇到了许多实际的和方法学上的问题。这些问题可以归结如下:

#### ● 系统的边界

将产品生命周期分解成各个过程时,经常遇到的问题,是究竟过程要分解到何种程度才算好?例如,在聚乙烯的生产中,首先要开采石油。石油要装在储油罐中运输。储油罐又要用钢材制造;因而生产钢时,便需先开采原料矿石。于是在生产石油和储油罐时,就要有采油的设备、炼油的设备、采矿的设备、炼钢的设备、轧钢的设备,乃至卷板机、焊接机等等。这样将使生命周期的过程树变得十分复杂。为了简化清单的编制,必须划定系统的边界。比如,通常规定生产用的设备一概不包括在清单中。

#### ● 生成一种以上产品的过程

以电解盐生产氯气为例。这时,电解过程的环境影响后果,便不能单独认为全部归于氯气。因为与此同时,还产生苛性钠和氢气;所以此处还需要有适当的分配规则。例如,按产品的质量或产品的经济价值来分配。

#### ● 被避免的各种环境影响

当一种处置过程能生成有利可图的输出时,例如在一家城市废弃物焚烧厂,虽然它会引起各种环境影响,但是它却能产生能源。按常规处置方式,并不需要生产能源或材料;但是焚烧的处置方式,至少能够减小环境影响。

为了能减少环境影响,因此便导入被避免的环境影响的思想。其实,被避免的各种环境影响,也就是在材料和能源的实际生产中会出现的各种环境影响的当量。某种过程被避免的环境影响,是从其他过程所引起的种种环境影响推导出来的。

#### ● 地理上的变化

在所用电力对环境的有害性上,瑞典的一家电解厂比荷兰同样一家电解厂要小。这是因为瑞典所用的,是极为丰富的水力发电。这就说明,同样的工业生产过程,由于地理位置的不同,会产生绝然不同的环境影响。

### ● 数据质量

有关环境数据方面的出版物,往往是不完整的或不正确的;而且有些数据已经过时,在过去 10 多年中,加工工业已将排放物减少约 90%,而许多场合下,不少出版物却仍然保留着过时的数据。显然,沿用过时的数据将会导致错误。

### ● 技术选择

人们可以在最差(落后),一般和最好(先进)的技术之间作出区分。在开始收集数据前,重要的是要弄清楚,你对哪种类型的技术感兴趣。通常,都是尽可能地收集一般常用技术,作为编制环境影响因素清单的基础。

以上这些问题,是在编制环境影响因素清单前,首先要明确和解决的。除去这些问题外,通常较为可行的是,编制一份环境影响清单。然而,不太合理的是,将其结果当作一种绝对真理。在解释那些环境影响时,必须考虑诸如上述技术选择、系统边界、数据质量等因素。这也正是为什么在评定有关产品的环境完美无缺方面,专家之间似乎存在着不同意见的原因。

### (3) 环境影响的评估

环境影响因素清单是 LCA 研究最为客观的一种结果。然而,一份造成环境影响后果的物质的清单是很难解释清楚的。为了使得这项解释任务变得更为简洁,于是便采用产品生命周期的环境影响评估方法来评价这些环境影响。

环境影响评估中存在着两个问题:

- 尚无充足的数据去计算由一种环境影响所引起的生态系统损害。
- 即使生态系统的损害是可以计算的,但是也还没有一种被普遍接受的方法,去评估对生态系统的损害。

最早的环境影响评估方法之一,便是瑞典一家研究机构开发的“环境优先策略”(EPS; Environmental Priority Strategy)系统。在这一方法中,对每一种环境影响的完整因果链,按对人的当量,进行了计算。

另外一种方法是为瑞士政府开发的所谓“生态计分法”(Ecopoint)。它是基于“距离-到-目标”的原理。此处假设一种环境影响的当前水平和目标水平之间的距离,代表了排放物的严重程度。

这里我们介绍的是 SETAC 和 CML 开发的环境影响评估一般方法。它是通过计算环境影响后果来评估的。这一方法包含三个步骤,兹特分述于后:

### ● 分类与特征化

在分类步骤中,所有的物质根据其对环境产生的影响后果面划分为类。例如造成暖室效应的物质,或造成臭氧层空洞的物质,这样便划分为两类。某些物质可

能同时包含在几个类内。例如,如图 37.2-13 所示,NO<sub>x</sub>可以在有毒、酸化以及富营养化等三个类中出现。目前,在这一评估方法中,共计划分为 8 类,即如上一小节讨论生态指标法时已看到的:

- 暖室效应:主要是向大气排放大量 CO<sub>2</sub> 废气所造成
- 酸化:向大气排放大量 SO<sub>2</sub> 和 NO<sub>x</sub> 废气所致
- 臭氧层空洞化:指冰箱、空调等传统制冷设备中所用的氟利昂(氯氟碳化合物)
- 富营养化:是由 SO<sub>2</sub>、NO<sub>x</sub> 和磷引起
- 重金属:指原子序数大于其离子,且密度大于 5.0g/cm<sup>3</sup> 的金属。许多重金属如铅、镉、铬、汞等,都具有较大毒性。其排放物渗入地下水或排入河流,均将污染水质,危害人类健康。
- 致癌物质
- 夏季烟雾
- 冬季烟雾

在每一类范围内,许多物质是组合在一起而产生一个环境影响后果得分。这时,由于只是把涉及这一环境影响后果的所有物质的各自数量加起来,而并未施加权重,因此这样的处理是很不充分的。因为有些物质可能比其他的物质产生更为强烈的影响后果。对于这个问题,主要是通过通过对不同的物质施加不同的加权因子来解决。这一步骤便称为特征化步骤。

以前面的图 37.2-9 为例,解释图上的各种环境影响后果的得分,或许比解释造成这些环境影响后果的有关物质的清单更容易。但是,这种解释是无意义的。如果对一种产品而言,它的所有环境影响后果得分,都比另一种产品的得分要高,则就很容易作出哪个产品对环境影响更小的结论。如今的问题在于,这种产品的酸化得分较高,而另一种产品则是暖室效应得分较高。这时,你就很难对这种情况作出孰好孰坏的判断。

要对环境影响后果作出正确的解释,这就取决于两个因素:

- 与其他环境影响后果的大小相比较后的某一环境影响后果的相对大小。在图 37.2-9 的这个例子中,重要的是去看,致癌物质的得分 100% 是否是参照一个很高的或极低的环境影响后果水平。这就是规范化。
- 附在各种环境影响后果上的相对重要性。这就是评价。

### ● 规范化

为了获得对一种环境影响后果的相对大小有更好的了解,需要有一种规范化的步骤。为一种产品计算其每一种环境影响后果,其实是针对已知该类总的环境影响后果所作的标准测试。例如,生态指标便是用平均



一个欧洲人一年所造成的各种环境影响后果去作规范化的。当然你也可以为规范化选择其他基础。

规范化使你能够看到从材料生产到每一种业已存在的环境影响后果的相对贡献。图 37.2-10 便是规范化的结果。

● 规范化后的环境影响后果得分的评价

规范化大大地改善了我们对得分结果的判断能力。然而，我们依旧无法作出最终的评判。这是因为并

非所有的环境影响后果都能看成是同等重要的。在这评价阶段，规范化后的环境影响后果得分，都要乘以一个代表该环境影响后果相对重要性的权重，如图 37.2-11 所示。经过如此处理后，所有的得分便可直接相加求和，如图 37.2-12 所示。因此便可得到最终结果；而这一结果便能直观地显示出不同设计方案环境影响的优劣。

## 第3章 绿色产品设计中的设计准则

正如上一章中提到的,绿色产品设计方法并非与传统产品设计方法完全对立,而另起炉灶、自己单独搞一套。它恰恰是在继承传统产品设计方法的基础上,补充了有关绿色设计的内容和要求。这时所采取的补充方式,主要是提出了针对产品生命周期各阶段的相关设计准则,借此指导设计人员在产品设计过程中恪守准则而确保符合绿色产品的要求。

在传统产品设计时也有许多诸如零件材料选择、零件联接型式选择等准则。因此,凡是传统产品设计中已有的,便不再重复。此处所给出的设计准则,主要是绿色产品设计所特有的,并紧紧围绕着绿色产品设计最根本的目标:提高资源的利用率和对环境污染的最小化。

### 3.1 绿色产品设计中的结构设计准则

绿色产品设计中的产品结构设计是节约资源的主要方面。好的产品结构设计,既能减轻产品重量、节约大量材料和能源,又便于加工、装配、缩短生产周期,降低生命周期成本。为此,在设计时,应认真贯彻以下产品结构设计准则。

#### (1) 注意产品的节能省料

这里所强调的,不只是指在产品本身制造(即产品生命周期中的制造阶段)时,要节约能源和节省材料,而且还包括产品在使用阶段、维修阶段乃至最终处置阶段,都应考虑最大限度地节约能源和节省材料。惟其如此,才能全面体现绿色产品设计所奉行的节约资源和减轻环境负担的根本目的。

#### (2) 尽量减少构成产品的零件数

这一准则不仅要求设计人员在设计产品时,力求使相同功能的产品结构,其所需零件数应最少,从而可节省制造产品所需材料,达到节约资源的目的;与此同时,减少零件数也将相应地简化产品的生产计划与管理工作,简化产品的装配作业。因此,贯彻这一准则将可显著降低产品的成本。在产品生命终止后,由于构成产品的零件数为最少,故而将减少回收后的产品拆卸工作量,也即将降低拆卸成本。

#### (3) 尽量采用模块化设计

产品结构的模块化,将有利于产品的维护、升级和重用。因为模块化结构易于查找产品的故障,在找出有故障的模块后,即可迅速按模块更换,所以它能节约排除故障的时间和工作量,从而降低产品的维护费用。当

技术进步而促使产品更新换代时,模块化结构提供了用技术含量高的新模块替换旧模块的可能性,使产品能迅速而经济地实现升级。此外,因故障而被置换的模块,经过修理和翻新,可以作为维修用备件而被重用。这比用全新的模块作维修备件要经济合理得多。这是因为产品需要维修时,这时产品本身已经不是全新的产品,充其量也不过只有几成新而已,因此无需用全新的模块去替换;其次,全新的模块成本必然要比翻新的模块成本高许多,所以用翻新的模块去维修既经济又合理。

#### (4) 应将价值高的零部件布置在易于分离的部位

此处所谓的高价值零部件,显然是指其回收价值高。这类零部件往往是由价值高的材料,以及极为复杂的工艺过程所制成的,所以值得回收。为了便于回收,在设计时,应将它们布置在易于拆卸的部位,以便能迅速拆下,借此提高拆卸的回报率。

#### (5) 应将有毒或有害材料制成的零部件布置在易于分离的部位

为了最大限度地避免在拆卸时产生不必要的环境和人体的负面影响,用有毒或有害材料制成的零部件,在设计时应该尽可能地将这类零部件布置在便于迅速而安全拆离的部位。

#### (6) 尽量将无法回收的零部件集中布置

对于无法回收利用的零部件,在产品的设计时,应该将它们集中布置。这样做,是为了减少拆卸时间,提高拆卸效率,降低拆卸成本。既然这些零部件属于无法回收,因此就不必予以拆卸,可直接将这些零部件粉碎、分选、打包、回炉,而作材料回收。

#### (7) 尽量避免采用镶嵌结构,如塑料零件中嵌入金属件

为了回收某些塑料,这类塑料中便不宜嵌入各种金属件。因为如果在塑料中嵌入金属件,这将增加先行从塑料中拆除金属件的附加工序,从而增加回收成本,否则便无法进行粉碎作业。此外,若不取出金属件,还将影响材料回收的纯度和未来材料的品质。

#### (8) 应尽量保证产品被拆卸时稳固牢靠

在产品的设计时,应考虑产品日后被拆卸时能有一稳固牢靠的基础件。这样即可不必采用专用的拆卸夹具,便进行高效率的手工拆卸,从而显著降低拆卸费用。

#### (9) 尽量减少接线数量和接线长度

许多产品的电气部分都会用到各种电线或电缆。由于电线和电缆较柔软,在拆卸时较为不便,且无端多用也纯属浪费,因此应尽可能减少接线数量和接线长度。

### 3.2 绿色产品设计中的材料选择准则

产品设计阶段中的重要问题,除了产品总体结构设计必须满足其功能要求外,其次就是在详细设计阶段对零件材料的选择。根据绿色产品设计的要求,在选择零件的材料时,应该遵守下列准则:

#### (1) 尽量减少产品中所用材料的品种数

减少产品中零部件制造时所用的材料品种数量,不仅简化材料的采购工作和仓库的管理工作,从而可减少企业的经营管理成本,同时也为日后产品生命终止而回收时的拆卸工作带来极大的方便。因为当废弃产品中的零部件作为材料来回收时,这一准则将显著地简化材料的分选工作。材料品种数量愈多,则材料分选工作愈复杂,花费在零部件拆卸和按材料品种分选的时间也愈多。显而易见,这将大大增加回收成本。试设想,如果一个部件中的零件是由同一种材料制成,则此时该部件不必拆卸和分选,即可整体粉碎、打包、回炉。

#### (2) 尽可能使相互联接的零部件材料能兼容

同理,在考虑零部件作材料回收时,这时相互联接的零部件所用的材料若是相同或兼容,则可不必修卸和分选,便能直接粉碎、打包、回炉。这样便可免去将零部件拆卸和分选的作业,因此可以节省拆卸和分选的成本,从而降低产品生命终止后的处置费用。

#### (3) 尽量选用可回收或易回收的材料

随着人们环境意识的不断提高,人们逐渐感到发展循环型经济的重要性。只有确立这样一种经济模式,才能确保人类社会与经济的可持续发展。循环型经济的本质,即生产一个产品所用的资源,是可以不断循环利用的。比如,首次生产某个产品所消耗的资源为100%,待到产品生命终止而被废弃不用时,此时将该废弃产品回收回来,其资源的95%可以重新利用,因此只需补充5%的新资源,便可重新生产一个新产品。为了达到重复利用或循环使用资源的目的,设计人员必须尽可能选用易于回收或可以回收的材料。在目前的生产技术条件下,金属材料的可回收性较好,非金属的复合材料的可回收性较差;塑料中的热塑性塑料可回收重复利用,而热固性塑料的分解与回收均较困难;单一材料的回收技术较为成熟,但复合材料的回收仍较困难。此处按材料可回收难易程度,提供一张选用表(表37.3-1),以资参考。

表 37.3-1 常用材料的可回收难易程度

回收性能好	回收性能中等	回收性能差
贵金属:金、银、铂、钯 非铁金属:锡、铜、铝合金 铁金属:钢及其它合金钢	非铁金属:黄铜、镍 塑料:热塑性塑料 非金属:纸制品、玻璃	非铁金属:铅、锌 塑料:热固性塑料 非金属:陶瓷、橡胶 其他:用不同工艺方法使两种材料复合在一起者,如涂层、镀层、铆接、粘接、镶嵌等

由表中可知,这里将材料可回收的难易程度划分为好、中、差三类。因此在选用材料时,应优先选用回收性能好的材料;而尽量少用或不用回收性能差的材料。这对提高资源利用率,减轻环境负担有着重要意义。

#### (4) 对材料进行必要的标识,尤其是塑料类材料

对于材料进行应有的标识,是很有必要的。这样将使产品生命终止后的回收处置工作大为简化。可以节省大量在识别与分选零件材料时的工作量,从而显著提高回收效率,降低回收成本。

特别是塑料的回收,愈见其重要性。因为塑料以其价格低廉、质轻、耐磨、抗腐蚀、绝缘性能好、易于成型等特点,所以半个多世纪以来,备受设计人员的青睐。通过在塑料中增添各种添加剂和增塑剂,即可得到满足各种不同物理化学的和机械的性能,从而使塑料广泛地替代常用金属材料,以致在产品 and 工程中占有越来越重要的地位。常用的塑料品种有:聚乙烯(PE)、聚氯乙烯(PVC)、聚碳酸酯(PC)、聚丙烯(PP)、聚苯乙烯(PS)、ABS树脂等。它们主要用来制

造各种模塑制品、泡沫塑料、绝缘材料、有机玻璃、层压板材、涂敷材料、粘结剂、管道、薄膜、纤维、织物、胶片等。正因为塑料被广泛应用,所以塑料制品在城市垃圾中的体积比例已超过20%;在填埋处置的废弃物中,其体积比例也已超过18%。这些比例在未来还会继续上升。如果所有塑料制品均可重用,则据测算,被废弃的塑料总量将可减少50%。由此可见,塑料这类材料的可回收性,是绿色产品设计中所应考虑的主要环境属性。

#### (5) 尽量选用回收后的再生(二次)材料

将废弃产品回收回来,经过拆卸、粉碎、分选、打包、回炉,重新冶炼成二次材料,目的在于重新进入产品生产过程,而重新生产出新一轮的产品。因此,设计人员应该优先选用二次材料。否则,重新冶炼成二次材料便无任何意义;而大声呼唤节约资源、贯彻可持续发展战略,将大半成为空话。国外也曾经遭到企业拒用二次材料的尴尬。因为二次材料有时在性能和品质上总要比一次材料稍差一些,毕竟已被使用过一次而重新

回炉冶炼的。为了鼓励企业积极采用二次材料,国外政府主要运用政策进行调控,通过经济手段给积极采用二次材料的企业以经济上的实惠,借此促进二次材料的广泛使用。

(6) 粘贴的标签材料或涂刷的标记材料应尽量与基体材料兼容

设置这一准则的出发点,主要是为了在回收材料时,可以不必清除标签和标记。实际上,许多标签的材料或标记的涂料,往往是与被粘贴的或被涂刷的基体材料完全不同。因而在回收基体材料时,就必须将标签或标记清除干净,否则将影响未来所回收基体材料的纯度。清除标签或标记的作业,是一件既费时又费力的作业,所以这项清除作业的成本很高。若能使基体材料与标签标记的用料兼容或一致,则既可确保所回收基体材料的纯度,又可节省原本需要清除标签或标记的高昂费用。

(7) 尽可能不用或少用有毒或有害的材料

不少材料在加工过程中会产生对环境和人体的不利影响。这是因为有些材料本身有毒或有害,而有些材料可能在加工或处理时会排放有毒和有害排放物。所以在设计产品时,尽量避免选用这类材料。

有毒或有害材料的生产或制备、存储、使用、回收和处置的过程,均极为复杂而难度大,且费用十分昂贵。为此,一方面必须对这类材料的废弃物进行严格回收和安全处置;另一方面要在产品设计选择材料时进行慎重考虑和仔细筛选。下面列出一些过去常用的有毒有害材料,供选用材料时倍加注意:

- **重金属** 常用的重金属有铜、镍、锌、铍、铅、镉、铊、铬等。目前对于铜、镍、锌的回收和处置,都已有较为有效的技术和方法,但对于铍、铅、镉、铊等则尚无成熟的技术和方法。铍主要用于宇航工业中各种紧固件的生产。采用铍的工厂需经特殊设计,工厂应设置在与外界环境相对隔离的地区,以防止含铍的物质扩散或释放到其他环境中去。镉作为塑料的稳定剂、油漆中的着色剂、与镍在一起构成可充电电池等有着较广泛的应用。然而镉是一种有毒物质,仅需 1g 镉便构成致命剂量;因此在产品设计选用材料时必须慎用。铬与镍均属合金钢中的重要合金元素。此外,铬在电镀中也用得很多。六价铬是一种致癌物质。如果一旦接触了皮肤,将使皮肤烧伤或腐蚀。因此,在镀铬过程中,务必采取相应的措施,将镀铬零部件上残留的六价铬,转化为无害的三价铬。铅在轴瓦、钎焊焊料、蓄电池和显象管玻壳中均有应用。为避免铅中毒,故尽可能不用或少用铅。

目前已研究出不含铅的代用材料,如不含铅的焊接用焊料。

- **石棉** 白色石棉已被证明是一种危险的致癌物质。过去白色石棉常用作隔热阻燃的重要材料。如今已有代用品,因此,不宜再选用白色石棉。虽然除白色石棉外,还有蓝色石棉和 X 色石棉。这后两种石棉虽无大害,但生产它们的代价太大,故迄今尚不具备实用性。
- **汞(水银)** 汞常用于钮扣电池、灯管、气压计、温度计、其他测量计、活性碳过滤设备等中。由于这些含汞的产品损坏或废弃后,未能严加管制和回收,任其随意丢弃,结果严重污染水资源,包括河流和湖泊。因此,对汞的选用也须慎之又慎。
- **氟化物** 氟化物的盐类是机械制造业对金属材料进行热处理渗氮时主要渗氮-传热介质。氟化物是剧毒化学品。尽量不用或必须慎用;且在不得已而选用时,则应严加保管与控制。
- **氟碳烃类(CFC)和有机卤化物** 传统的制冷剂氟利昂,即属于氟碳烃类化合物。许多国家已将氟碳烃类和有机卤化物定为有害物质。它们是破坏保护地球不受太阳紫外线直射侵害的臭氧层的罪魁祸首。目前已经开发出能够取代氟利昂的新型无害制冷剂。根据“蒙特利尔议定书”,我国也已将氟利昂列为禁用制冷物质。

(8) 必须使用有毒或有害材料时,应作标记

凡属功能或结构上必须采用有毒或有害材料时,此时为保证安全起见,必须在采用这类材料制成的零部件上,作出相应醒目的标记,以致识别和引起注意。避免未来在回收拆卸时,发生不必要的危及环境和人身安全的事故。

(9) 尽量避免采用粘结材料,除非被粘结件材料与之兼容

与此准则相类似的,已在上面叙述过,故此处便不再赘言。

### 3.3 绿色产品设计中的工艺选择准则

在绿色产品设计中,研究尚不充分的,当属绿色工艺这一块。这从已经发表和出版的有关绿色制造的论文和书籍中,即可知其一二。此处仅将散见于文献资料中所能看到的、有关工艺选择的一鳞半爪,汇集于此,以资参考。

(1) 尽可能使毛坯的尺寸与形状接近于其零件的最终尺寸与形状

遵守这一准则就能减少加工余量。因此便可节约

材料、有效提高材料的利用率。反之,若毛坯的尺寸和形状与零件的差别过大,则加工余量也必然过大,因此原本有用的材料经切削加工而变成无用的切屑,显然这不仅浪费了材料,而且还白白浪费了由有用材料转变的无用切屑时的能源和刀具。由此可见,遵守这一准则,对有效而综合地节约资源,有着积极而重要的意义。为此,必须采用已有的、并进一步开发各种能满足少切屑或无切屑的毛坯制造方法。

(2) 零件切削加工时,应尽量不用或少用冷却润滑液

为了减轻切削加工过程中工件与刀具之间的摩擦,以及避免由此所产生的切削热,因此在切削加工时往往采用大量冷却润滑液。冷却润滑液会老化而失效,为此需经常置换更新。废弃的冷却润滑液必须经过处理方可排放。显然,在制备新的冷却润滑液和处理废弃的冷却润滑液,以及在使用冷却润滑液时,均将消耗大量资源。甚至在美国,还出现处理冷却润滑液的费用远比制备冷却润滑液的费用要高得多的情况。因此,在产品的结构设计和材料选用时,应考虑未来零件的切削加工能不用或少用冷却润滑液。尽量采用干切削和少无切屑等新的工艺方法。这将减轻冷却润滑液对环境造成的负担,同时也不使操作者的健康因长期接触冷却润滑液而受到伤害。

(3) 尽量采用各种节能省料而又无污染的工艺

产品由毛坯加工成零件,进而装配成部件和最终商品的生产过程中,消耗着可以通过成本核算而看得见的大量材料、能源和人力。因此人们自有史以来,便始终孜孜不倦地研究着,如何使生产过程能以最小投入而获得最大产出。过去只是单纯考虑生产本身的效益问题;如今则必须同时结合环境负担问题来考虑。由此可知,今天人们所希望的各种新的工艺方法,使不止是能节能省料而已,而且还必须是无污染的,或者说是对环境友好的。

### 3.4 绿色产品设计中的产品使用阶段的设计准则

在传统产品设计中,设计人员主要把注意力集中在考虑解决满足产品的功能上,至于产品在使用期间是否对生态环境和人体健康造成危害,则往往考虑极少。此外,设计人员对于产品成本的考虑,也仅仅局限于产品本身开发、设计和制造的成本而已。对于产品使用期间所必须支出的经常性运作成本,也很少替顾客作深入考虑。以致不少产品在顾客眼中变成了“买得起而用不起”的产品。为此,在产品使用阶段便须把设计的注意力,从单纯的产品功能设计,延伸到产品使用阶段的环境影响和使用成本上来。

(1) 应使所设计的产品在使用阶段是节能的

正如上面所说的,不仅产品在其生产阶段是节能的,而且在其使用阶段也应是节能的。在产品的整个生命周期中,产品的生产阶段所占的时间比例是极小的,而产品的使用阶段却占据了时间比例的绝大部分。这说明节能效果的重点是在产品的使用阶段。因此,在产品的概念设计时,便需认真从产品的工作原理上去研究和解决产品使用阶段的节能问题。要把满足产品的基本功能与其使用过程中的节能要求结合起来,而非割裂开来研究。

(2) 尽可能减少产品使用过程中所排放的各种(气体、液体或固体)废弃物

减少产品使用阶段所排放的各种废弃物,是产品最重要的环境属性。大多数废弃物都会污染环境;或者破坏生态环境,或者损害人体健康。一种产品对环境所构成的最大可能的污染或伤害,也主要集中在使用阶段。因此,设计人员在设计产品时,应该将大部分注意力倾注于产品使用阶段的环境问题。

(3) 保证所设计产品的可靠性,使之经久耐用

提高产品的可靠性,提高产品的使用寿命,实际上也是一种积极的节约资源的途径。因为如果使一种产品的使用寿命比原来产品的使用寿命延长了一倍,这就意味着这种产品一台相当于原来产品的两台。这样岂非无形中就节约了生产整整一台产品的资源?

(4) 避免产品使用过程中可能产生的振动和噪声

### 3.5 绿色产品设计中产品维修的设计准则

顾客在购买产品后,除了要承担经常性的运作费用外,还要在其使用寿命期间支付必要的产品维护与修理费用。如何节约这笔维修费用,这便涉及到设计人员在设计产品结构时,应该考虑产品维修性准则的问题:

(1) 充分考虑零部件维修时的可达性

为了减轻维修的工作量,缩短维修的时间,应该在结构设计时,充分考虑维修作业的可达性。为此,应该设计必要的手孔或人孔,使维修人员的手或人可由此直接达到所需维修处,而不致于要拆卸许多其他零部件方可达到。

(2) 优先选用标准化的零件

产品结构时,若能优先选用标准化零件,则将大大节约维修成本。因为标准化零件可互换,而且易于迅速从市场上获得,所以无需自行专门制造维修用的备件。

(3) 优先考虑模块化的部件设计

产品的结构采用模块化部件拼装的设计思想,将有利于日后的故障诊断和维修。因为每一模块都具有它固有的独立功能,所以一旦产品发生故障,某一功能

失效,便能立即诊断并找出提供该功能的故障模块。为了不使故障引起太长的中断,这时可用同样功能的备用模块作迅速替换,从而可以大大节约故障现场的诊断和维修时间,迅速恢复产品的正常运行,减少因故障中断而造成的停机时间损失。

(4) 按照经久耐用的长寿命原则设计产品

如上所述,经久耐用的长寿命产品,不仅延长了产品的使用期限,同时也减少了产品的维修次数。减少维修次数,意味着减少相应的维修费用,其中包括减少了维修时所消耗的资源。

(5) 凡易于发生装配差错处,应作防差错设计

(6) 在易磨损的机构中,考虑采用可调节和补偿的结构

对于产品工作中易磨损的零件,为了不必频繁更换这种易磨损零件,以节约制造这种易磨损零件的资源,简化维修的工作量,因此可以采用可调节和补偿磨损的结构。

### 3.6 绿色产品设计中产品回收阶段的设计准则

传统产品设计与绿色产品设计的最大差别之一是,前者在设计阶段并不刻意去考虑产品寿命终止后

的回收和再生利用问题;而后者为了资源的循环利用、减轻环境负担,就必须认真考虑产品寿命终止后的回收和再生利用问题。这从第 2 章的图 37.2-2 与图 37.2-3 中,即可看出二者的这一差别。因此,在进行绿色产品设计时,必须使产品结构充分设计充分考虑产品的可回收性。

废弃产品的最终处置必须通过回收来实现。回收后的处置决策,除了个别情况可将整台产品回收后作二手产品重用外,大多数废弃产品是按图 37.3-1 所示的处置策略。图中加阴影的长方形框表示产品生命周期主要阶段的活动;椭圆形框则分别表示相应的处置策略,即:零件直接利用、零部件再制造、材料再循环以及无法回收利用的资源只能焚烧或掩埋。前三者的指向向上的箭头,则分别表示各自回收处置后的归属。针对不同的回收处置策略,其回收设计的准则也不同。

当考虑废弃产品中的零部件需重用时,则在废弃产品回收后应对其进行拆卸和分解。为此,设计人员有必要了解产品拆卸分解的过程,以便在设计产品时能充分考虑拆卸和分解的要求。

当考虑废弃产品作为材料再循环时,这时会出现两种情况:

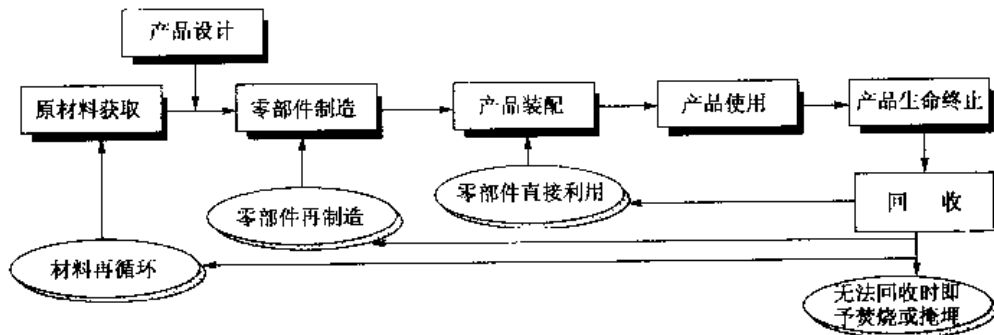


图 37.3-1 废弃产品回收后的四种处置策略

● 手工拆卸 手工将废弃产品拆卸分解成零件,然后按零件材料分选成单一材料回炉。

● 机器粉碎 废弃产品不经拆卸分解,直接用粉碎机将其粉碎,然后由分选机分选成单一材料回炉。

废弃产品一般的拆卸和分解过程如图 37.3-2 所示。回收设计的准则主要是针对这一过程来考虑的。

产品设计必须考虑解决的基本问题为:产品的总体结构、零部件的结构、零部件的材料和零部件的联接方法;然而这四个方面对于废弃产品的最终处置策略,却有着直接的影响。在对废弃产品最终处置策略作出选择时,固然主要的依据是回收的经济性,可是本质上,回收的经济性仍然取决于上述这四个基本因素。例如,当重用废弃产品中的零部件成为不可行时,可能主

要的原因是拆卸的费用过高,或是难以拆卸,或者拆卸时极易损坏零部件而导致无法重用。而拆卸费用过高、难以拆卸、或拆卸时易损坏等,归根结底,是产品的总体结构,或零部件的结构,乃至零部件的联接方法不合理所致。因此,在考虑回收的设计准则时,除了要考虑废弃产品的拆卸过程外,还应该考虑产品的结构设计要便于拆卸。下面即分别阐述有关设计准则。

(1) 满足拆卸要求的可拆卸性设计准则

● 产品的总体结构划分清晰 这是指产品可以划分成明显的独立部件,因此拆卸和分解废弃产品时,便易于按部件有条不紊地进行。

● 零部件的联接处迅速可达 零件与零件间的联接,以及联接与部件间的联接,应设计成迅速可达。

这样便可迅速进行拆卸和分解,因而大大提高拆卸和分解的效率。

- **联接件的数量应最少** 显然,联接件的数量越多,则拆卸分解越费时,拆卸费用也越高。因此,应在产品设计时,使所用的联接件数量为最少。以此节省拆卸的时间,降低拆卸费用

- **联接件应采用标准件** 采用标准的联接件不仅能直接降低产品成本,同时也便于回收拆卸。因为这时可以沿用标准的拆卸工具,而无需配置专用工具,所以也可节约拆卸费用。

- **减少拆卸用的工具种类** 减少拆卸用的工具种类,便可使拆卸过程中不必频繁更换拆卸工具,从而可提高拆卸工作效率。为此,在产品设计时,便应限制紧固件的种类。

- **拆卸时应不使零部件易受损** 在产品结构设计时,既要考虑便于零部件的拆卸,又要考虑拆卸时零部件不易受损,否则,拆卸下来的零部件便不复重新使用。

- **拆卸过程应尽可能短** 这就要求在产品设计时,尽量使产品和部件的结构简单。惟其结构简单,才能使拆卸过程短;拆卸过程越短,则拆卸效率也越高,拆卸费用也越低。

#### (2) 满足清洗要求的设计准则

- **应使清洗剂易于到达零件上的所有部位,并且清洗后又不易积存残液** 为了满足这一要求,在零件

结构设计时,尤需注意零件内腔及其转角处的结构形状。务必做到清洗剂既易达到,而清洗后又不留残液。否则,残留的清洗剂会腐蚀机件。

- **零件材料应对清洗剂和清洗技术有耐受性** 这在选用零件材料时应加注意。有些材料与清洗剂易起化学反应而导致材料变质、损伤或失效。有些材料不适合某种清洗技术,因此,在为零件选用材料时,必须考虑未来回收拆卸清洗时所用的是何种清洗剂和何种清洗技术,然后据以选择相应的耐受材料。

#### (3) 满足检测和分选要求的设计准则

- **应便于检测零件上易磨损和易生裂纹的部位** 为了迅速判别零件是否可以重用,必须对零件进行检测。这时,零件的结构应便于快速检测出其磨损和裂纹所在。因此,须使检测仪器的探头易于迅速而无阻碍地到达零件上的受检部位,以提高检测的效率,降低检测费用。

- **尽量采用标准化的零部件** 标准化的零部件将易于检测。这是因为标准化的零部件,经过广泛使用,结构较为成熟,易于产生磨损和裂纹的部位也较为明确,故而检测较容易。

- **当废弃产品作为材料回收利用而人工拆卸时,则零件上应有材料标记** 预先在零件上作出材料标记,则在拆卸分解出零件后,即可按零件上的材料标记分选成单一材料。这将大大提高分选的效率,并且不会将材料分错。

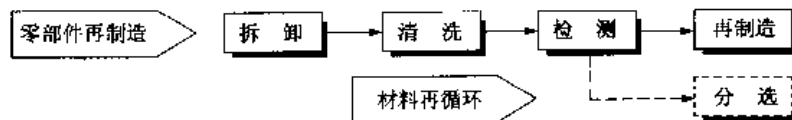


图 37.3.2 废弃产品回收后的拆卸分解过程

## 第4章 绿色产品设计的评价指标系统及其评价方法

### 4.1 绿色设计评价指标体系

有关绿色产品设计的评价值问题,其实已经在第2章的产品生命周期评价(LCA:Life Cycle Analysis)中作过详细阐述。国外所有的生态设计、面向环境的设计等,几乎都是以LCA作为评价自身对环境属性影响的基础,而LCA也已经有一套标准的分析评价框架和方法步骤。这可以参阅国际标准化组织(ISO)以及我国国家质量技术监督局发布的有关标准。

然而,LCA只是单纯的对环境属性影响的一个方面评价产品及其制造过程而已。固然对环境属性影响的评价,确实是绿色产品设计的关键所在,可是全面评价一个绿色产品设计,除了首先要考虑环境属性外,还必须同时考虑其技术属性和经济属性。片面强调或追求产品的环境属性,有时会导致忽视甚至牺牲其应有的技术属性和经济属性,这显然是不可取的。因为设计工作往往不可能一气呵成,而是要经过多次的评价、改进,才能达到真正的设计要求。绿色设计方案的评价又不是简单的单目标决策问题,而是个多目标决策问题,因此评价工作不能只针对产品的某一方面。

我们认为绿色设计评价指标体系必须系统地反映绿色产品的三大基本要素,即绿色产品的技术先进性、绿色产品的环境协调性和绿色产品的经济合理性。为此,我们提倡对绿色产品设计进行综合评价,因而提出一个综合评价指标系统。这一系统综合考虑了产品的环境属性、技术属性和经济属性。此处提出的绿色产品设计综合评价指标系统结构,即如图37.4-1所示。

由图中可知,上述绿色产品设计综合评价指标系统,可以分为四级。其中第一级即为绿色产品设计综合评价指标系统框架;第二级为绿色产品设计的三大基本属性,即技术属性、经济属性和环境属性;第三级是对第二级进行分类和细化,其中,技术属性分为功能性和技术先进性;经济属性分为生产成本、使用成本和处置成本;环境属性分为能源消耗、资源消耗、大气影响、水体影响、土地影响和人体健康等六小类。第四级是在第三级基础上,结合产品生命周期特性建立的较为具体的指标分类,如功能性由产品的基本功能和辅助功能组成,技术先进性则是由体现产品生命周期各阶段

的技术指标组成,如体现制造阶段的可制造性、使用阶段的安全性,维修和回收阶段的产品可拆卸性等。

### 4.2 绿色产品设计综合评价的方法

在综合评价绿色产品设计的办法上,通过对已有办法研究的分析,结合以上所提出的综合评价指标体系,此处拟采用层次分析法确定各级评价指标的权重;用加权平均法在对第四级项目指标进行打分的基础上,计算各级指标的分值;最终用模糊评价法对一级指标作出评价。以下将分别介绍层次分析法、加权平均法以及模糊评价法。

#### 4.2.1 层次分析法

层次分析法是一种实用的多准则决策方法。这种分析方法把一个复杂问题表示为有序的递阶层次结构。通过人们对待决策方案优劣的判断进行排序,并将决策中的定性和定量因素进行统一处理,因此方法具有简洁和系统等优点,很适合在复杂系统中使用。其基本步骤如下:

##### ● 构造层次结构模型

构造层次结构模型是层次分析法中的一个关键。一般,将问题的预定目标作为最高层的目标层,中间层次包括策略层、约束层、准则层等,最底层则为决策方案或指标。利用这些指标就可以简明地表示最高目标。必须指出,层次模型往往需要细分为多层,但过分细化也没有必要,细分层数必须根据具体问题而定。

##### ● 构造判断矩阵

在建立层次模型后,上下层次之间的元素隶属关系就已经确定。这样便可以比较下一层中与上一层所选定的某因素有关的各因素间的相对重要程度。层次分析法是通过因素间的两两对比来描述因素之间相对重要程度的,即每次比较只有两个因素,而衡量相对重要程度的差别是使用1-9比率标度法,即使用1、3、5、7、9或1/3、1/5、1/7、1/9来表示某因素相对于另一因素相对重要程度,具体含义如表37.4-1所示。

通过上述的两两比较,就得到进一步计算必需的判断矩阵。具体地说,假设A层因素 $A_i$ 与下一层即B层的因素 $B_1、B_2、\dots、B_n$ 有联系,判断矩阵便有如下形式:



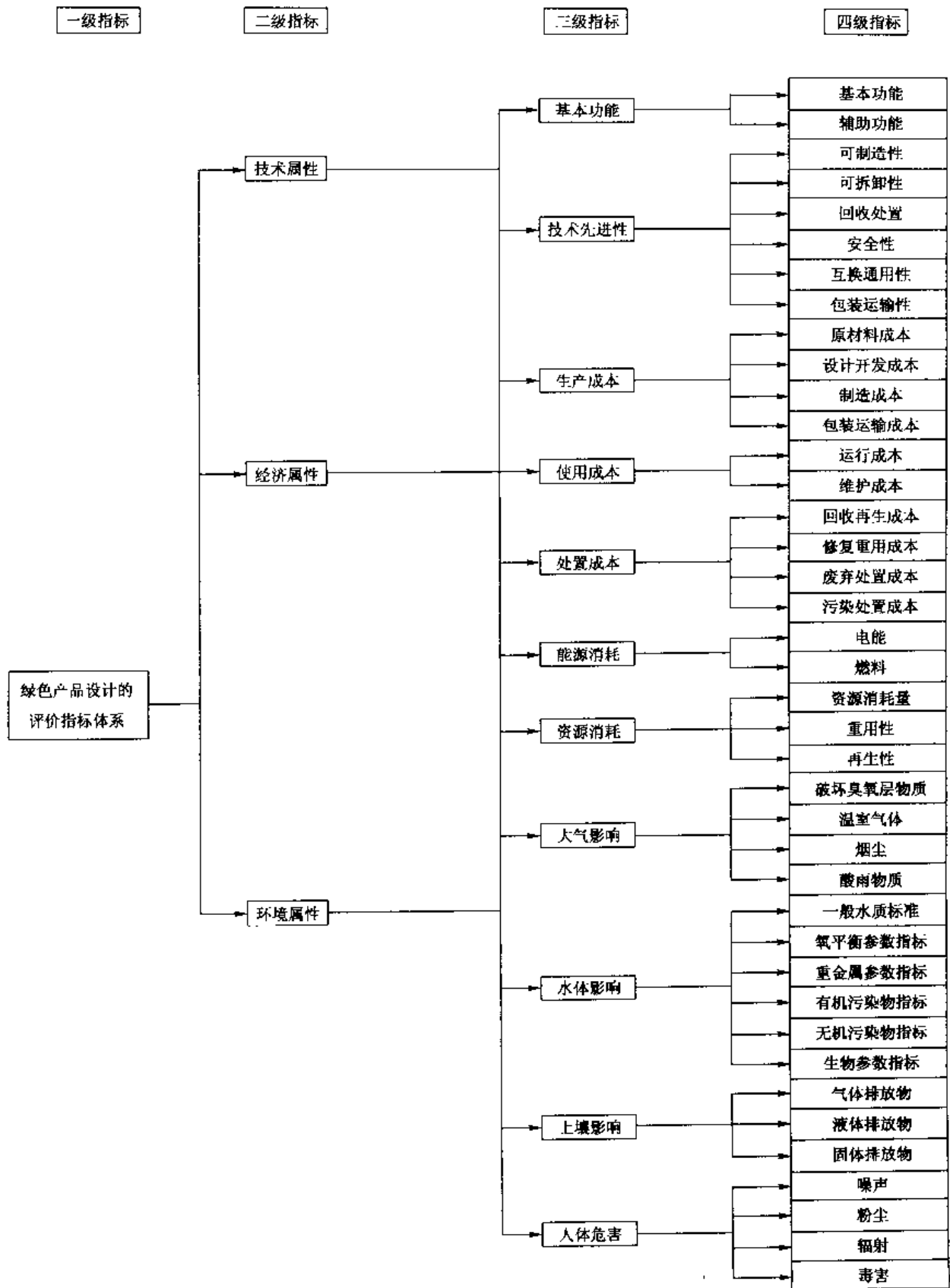


图 37.4-1 绿色产品设计综合评价系统结构

表 37.4-1 标度 1-9 的含义

标度	含 义
1	两因素相比, 具有同样重要性
3	两因素相比, 其中一个稍微重要些
5	两因素相比, 其中一个明显更重要
7	两因素相比, 其中一个强烈重要
9	两因素相比, 其中一个极端重要
倒数	若因素甲与因素乙相比的标度为 $i$ , 则乙比甲的重要标度为 $1/i$

$A_k$	$B_1$	$B_2$	...	$B_n$
$B_1$	$b_{11}$	$b_{12}$	...	$b_{1n}$
$B_2$	$b_{21}$	$b_{22}$	...	$b_{2n}$
...	...	...	...	...
$B_n$	$b_{n1}$	$b_{n2}$	...	$b_{nn}$

比如  $B_1$  比  $B_n$  的重要程度是  $b_{1n}$ , 这时判断矩阵记为  $A_k = (b_{ij})_{n \times n}$ . 判断矩阵元素的值反映了人们对各因素相对重要性的认识。如果所考虑的  $A$  层一共有  $K$  个因素, 那么  $k=1, 2, \dots, K$ . 因而这时就应该有  $K$  个判断矩阵, 并且必须一个一个地生成。

● 层次单准则排序及其一致性检验

实际上, 构造层次结构模型和判断矩阵的过程, 是一个信息分散以及对分散信息进行度量、表述的过程。判断矩阵实质上是对分散信息的一个完整描述和度量, 即将关于各因素重要程度差别的信息分散在判断矩阵的  $n \times n = n^2$  个元素中, 因此, 为了得到这些因素相对重要程度的信息, 必须对分散的判断矩阵的  $n^2$  个元素中的信息进行集中, 层次分析法采用的信息集中手段是特征向量法。

单准则排序是指根据判断矩阵而计算某一准则下各元素的相对权重, 并进行其一致性检验。设针对某一准则, 各元素的权重向量为  $W = (\omega_1, \omega_2, \omega_3, \dots, \omega_n)^T$ , 则可以通过求解下列方程得到  $W$

$$AW = \lambda_{\max} W \quad (37.4-1)$$

式中  $\lambda_{\max}$  为矩阵  $A$  的最大特征值。

由于矩阵  $A$  中的元素是通过主观判断确定的, 因此  $A$  不一定具有规范的一致性。通常采用幂乘法求  $\lambda_{\max}$  和  $W$  的值。计算步骤如下:

(1) 任意取一个正的规范化初始权重向量  $W_0$ 。如可令  $W_0 = (1/n \ 1/n \ 1/n)^T$ 。

(2) 依次计算  $\bar{W}^k = A W^{k-1}, k=1, 2, A$

(3) 令  $m = \sum_{i=1}^n \bar{\omega}_i^k$ , 计算  $W^k = \frac{1}{m} \bar{W}^k$

(4) 对于事先给定的计算精度  $\epsilon$ , 若满足  $|\omega_i^k - \omega_i^{k-1}| < \epsilon (i=1, 2, \dots, n)$  时, 则  $W = W_k$  即为所求的权重向量, 否则从步骤 (2) 开始继续计算。

(5) 计算矩阵  $A$  的最大特征值  $\lambda_{\max}$

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{a_i^k}{\omega_i^k} \quad (37.4-2)$$

计算  $W$  和  $\lambda_{\max}$  的幂乘法是通用的计算方法, 适合计算机计算。在精度要求不很高的情况下, 可采用几何平均法、规范列平均法等方法近似计算  $W$  和  $\lambda_{\max}$ 。

虽然在构造判断矩阵  $A$  时并不要求判断具有一致性, 但判断偏离一致性过大也是不允许的。因此需要对判断矩阵  $A$  进行一致性检验。主要一致性指标有:

■ 计算一致性指标  $CI$

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (37.4-3)$$

式中  $n$  为判断矩阵的阶数。

■ 计算相对一致性指标  $CR$

$$CR = \frac{CI}{RI} \quad (37.4-4)$$

式中  $RI$  为平均随机一致性指标, 是足够多个根据随机发生的判断矩阵计算的一致性指标的平均值。1-10 阶矩阵的  $RI$  取值如表 37.4-2 所示。

表 37.4-2 1-10 阶矩阵的  $RI$  取值

矩阵阶数	1	2	3	4	5	6	7	8	9	10
$RI$	0.0	0.0	0.5	0.90	1.1	1.2	1.3	1.4	1.4	1.4
	0	0	8	2	4	2	1	5	9	

一般而言  $CR$  愈小, 判断矩阵的一致性愈好, 通常认为  $CR \leq 0.1$  时, 判断矩阵具有满意的一致性。

● 层次综合排序

在上述计算每一层诸因素相对于上一层某一因素的单排序权重之后, 计算同一层诸因素相对于最高层的相对重要性排序权值, 就称为层次总体排序。这一过程是从最高层开始到最底层逐层进行的。具体计算公式为: 设上一层共有  $A_1, A_2, \dots, A_k$  个因素, 层次总排序之权值为  $a_1, a_2, \dots, a_k$ , 下一层共有  $B_1, B_2, \dots, B_n$  个因素, 各因素相对于  $A_1, A_2, \dots, A_k$  的单排序权值  $b_{ij}$ , 如表 37.4-3 所示。由表可看出  $B_i$  的层次总排序权值为:

$$b_i = \sum_{j=1}^k a_j b_{ij} \quad i=1, 2, A, n \quad (37.4-5)$$

当  $B_k$  与  $A_j$  无联系时,  $b_{kj}=0$ 。

表 37.4-3  $B$  层次总排序权重

层次 $B$	层次 $A$				$B$ 层次总排序权重
	$A_1$	$A_2$	...	$A_k$	
$B_1$	$a_1$	$a_2$	...	$a_k$	$\sum_{j=1}^k a_j b_{1j}$
$B_2$	$b_{21}$	$b_{22}$	...	$b_{2k}$	$\sum_{j=1}^k a_j b_{2j}$
...	...	...	...	...	...
$B_n$	$b_{n1}$	$b_{n2}$	...	$b_{nk}$	$\sum_{j=1}^k a_j b_{nj}$

### ● 层次总排序的一致性检验

在前述记号下, 设  $B$  层次某些因素对于  $A_j$  的单排序的一致性指标为  $CI$ , 平均随机一致性指标为  $RI$ , 则  $B$  层次总排序的相对一致性 (也称随机一致性比率) 为:

$$CR = \frac{\sum_{j=1}^k a_j CI_j}{\sum_{j=1}^k a_j RI_j} \quad (37.4-6)$$

当  $CR < 0.1$  时, 认为排序结果具有满意的一致性, 否则需要重新调整判断矩阵元素之值, 直到总排序结果具有满意的一致性为止。

上面分析了层次分析法解决问题的五个步骤。但值得指出的是, 在寿命周期分析的过程中, 层次分析法既有优越之处, 也有其局限性。其优越之处在于, 用层次分析法进行方案比选的全过程都是基于对指标或方案的两两比较, 在任何一个层次都不要求对产品的各评价指标作出定量描述。其局限性在于, 层次分析法只能得出评价指标体系中各指标的重要程度或备选方案相对优劣的排序结果。由于层次分析法并不能进行定量分析, 因此在绿色产品设计综合评价中, 主要用它来完成各评价指标权重的确定。

#### 4.2.2 加权平均法

加权平均法用于确定各层指标的分值, 可以用下面的式子来表示:

$$E_j = \sum_i a_i \cdot S_i$$

式中  $E_j$  ——第  $j$  个指标加权后的总分数;

$a_i$  ——第  $i$  个评价指标所占的权重, 且  $\sum a_i = 1$ ;

$S_i$  ——第  $i$  个评价指标的评分。

在加权平均法中, 权重的大小可以用层次分析法来获得, 而评分  $S_i$  则应根据具体的产品、行业标准等来确定。

对于最低层指标, 可以采用下面的方法确定:

#### ■ 专家打分法

专家打分法是一种简单易行的方法, 它多用于不可能在短时间内用统计方法、实验分析方法或因果论证方法得到评分的情况。目前它已广泛应用于经济学、环境学等科学技术领域中。然而, 该方法最大的缺点是其主观随意性较大。为了克服专家打分法的这一缺点, 建议在使用该方法时, 可向专家提出以下要求:

- 能够定量的指标尽量定量, 例如能够通过测量或计算得到的指标应尽可能量化;

- 单位相同的指标直接求和, 单位不同的指标必须进行无量纲化或转化为具有相同单位的指标, 例如: 为了衡量总的资源消耗量, 可通过建立资源与价格

之间的关系, 以某种材料价格为基准, 建立原材料之间的等价折算关系;

- 对于有国家标准和行业标准的指标, 则以标准为参考进行打分。对于没有国家标准的, 应尽量通过调研和数据收集获得打分依据;

- 不能量化和无章可循的指标应尽量避免; 实在无法避免时, 可利用专家知识进行打分。

#### ■ 模糊数学法

由于绿色设计中的评价指标有些具有一定的模糊性, 对于这些指标简单地用专家打分的方法往往缺乏说服力, 因此可利用模糊数学中的隶属函数和隶属度的概念来确定。

有关隶属函数和隶属度的定义如下:

设论域  $U$  上的一个模糊子集 (或模糊概念)  $\tilde{A}$ , 分别赋给  $\tilde{A}$  处于共维差异的中介过渡段的两个极点以 0 与 1 的数。在 0 到 1 的数轴上构成一个  $[0, 1]$  闭区间的连续统。对于任意  $u \in U$ , 都在连续统上指定了一个数  $u \tilde{A}(u)$ , 称为  $u$  对  $\tilde{A}$  的隶属度, 映射

$$\left. \begin{aligned} u \tilde{A}: U &\rightarrow [0, 1] \\ u &\mapsto u \tilde{A}(u) \in [0, 1] \end{aligned} \right\} \text{称为 } \tilde{A} \text{ 的隶属函数。}$$

通常, 隶属度的确定可采用统计法或隶属函数法。

#### ● 统计法

即收集一定量的评价信息通过统计得到隶属度。它尤其适用于没有一致评价标准的问题, 如对某机械产品外观因素的评价, 若其中 10% 的人评价为优, 评价为良和中的各 20%, 而有 50% 的人对其评价为差, 即可得出隶属评价集  $\{0.1, 0.2, 0.2, 0.5\}$ 。注意进行模糊统计试验次数应足够多, 以使统计得到的隶属度稳定在某一数值范围内。

#### ● 隶属函数法

隶属度也可通过隶属函数求得, 它一般适用于有一致评价趋势的问题, 如机械材料强度因素越高越好, 成本因素越低越好。例如: 开发成本  $C$  小于 10 万元为优, 30 万元为中, 大于 60 万元为差, 则成本为 50 万元的隶属度评价集可选直线型隶属函数, 如图 37.4.2 求得:

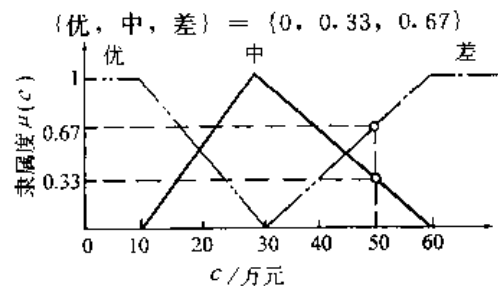


图 37.4-2 开发成本  $C$  的隶属度函数

### 4.2.3 模糊综合决策

由绿色设计评价指标框架可以看出,产品指标通常都表现出明显的层次性,而且部分指标具有一定的模糊性,因此可采用模糊综合评判方法进行综合评判。

模糊综合评判的方法就是将复杂的系统分解为比较简单的子系统,然后利用模糊数学对各子系统进行评估,再将其合成为对整个系统的综合评价。模糊综合评判能用较少的代价传递足够的信息,并能对复杂系统作出效率高和准确率高的判断和处理。其评判的主要过程如下:

- 根据产品的特点和欲达到的目标,构造产品评价指标的层次结构,即建立评价指标集  $U = (u_1, u_2, \dots, u_i, \dots, u_n)$ ;

- 针对某一指标,由评价者对该指标所支配的下一层元素进行模糊评价,建立评语集  $V = (v_1, v_2, \dots, v_j, \dots, v_m)$ ;

- 建立模糊评判矩阵;

给出模糊评判集的映射为:

$$f: U \rightarrow L(V) \quad (37.4-7)$$

$$f(u_i) = (r_{i1}, r_{i2}, \dots, r_{ij}, \dots, r_{im}) \in L(V) \quad i = 1, 2, \dots, m \quad (37.4-8)$$

式中  $L(V)$  表示  $V$  中所有模糊评语集;

$f(u_i)$  是关于因素  $u_i$  的评语模糊向量;

$r_{ij}$  表示关于  $u_i$  的具有评语  $v_j$  的程度。

由  $f$  导出  $U$  到  $V$  的模糊关系——模糊评判矩阵:

$$R = R_f = (r_{ij})_{n \times m}$$

- 利用层次分析法 (AHP) 建立权重集,即

$$A = (a_1, a_2, \dots, a_j, \dots, a_m), \text{ 且满足 } \sum_{j=1}^m a_j = 1;$$

- 进行模糊变换,并进行归一化处理;

模糊变换公式为:

$$B = A \circ R = (b_1, b_2, \dots, b_k, \dots, b_m) \quad (37.4-9)$$

其中:  $b_j = \bigvee_{k=1}^n (a_k \wedge p_{kj}), j = 1, 2, \dots, m$

归一化公式为:  $b_j = \frac{b_j}{\sum_{j=1}^m b_j}$

- 重复上述步骤,向上逐层合并,得到该产品的绿色度,供决策者参考。

### 4.3 实例验证

为了验证这一评价方法的实用性,此处采用对空调产品的绿色度评价作为实例。此实例的研究对象是某型号家用分体壁挂式空调器,其制冷量为 3500W,额定输入功率为 1350W,额定电流为 6.5A,室内噪声为 43dB,室外噪声为 52dB。

按以上所提出的绿色产品设计综合评价方法,层次分析法主要用于确定各三级指标对二级指标的权重,加权平均法用于计算各二级指标的评分,模糊评价法用于二级指标到一级指标的综合评分。综合指标系统的第四级则是按专家打分法给出。

在此实例中,仅以环境属性作为综合评价的对象。就方法论而言,技术属性和经济属性的综合评价的方法和过程,与此相同,故不在此赘言。

#### ■ 环境属性评价

首先为四级指标对二级指标的权重,如图 37.4-3 所示。此图实际上即为图 37.4-1 中二级指标层的环境属性部分,其三级指标分为能源消耗、资源消耗、大气影响、水体影响、固体废弃物影响、人体危害等六类。由此可知其判断矩阵分别为  $C_1, C_2, C_3, C_4, C_5, C_6$ 。

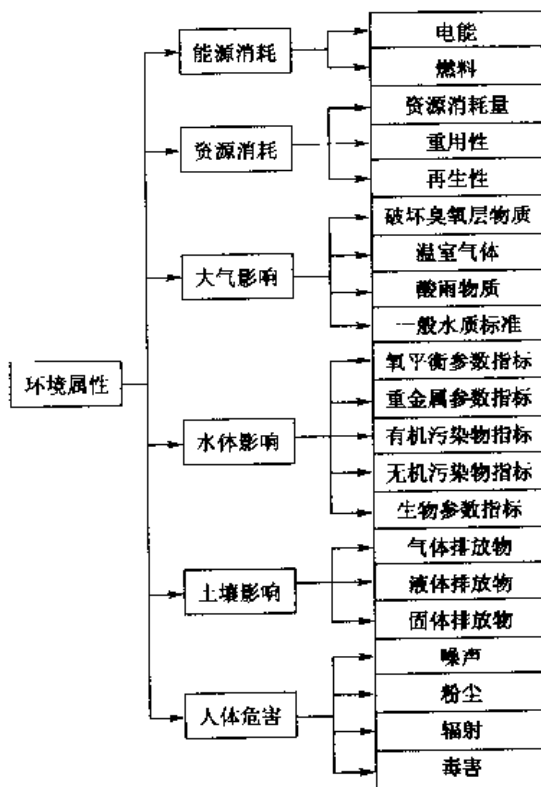


图 37.4-3 环境属性的指标子系统

根据层次分析法的方法及各指标之间的重要度,我们可以确定各判断矩阵为:

- 能源消耗

$$C_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- 资源消耗

$$C_2 = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{bmatrix}$$

● 大气影响

$$C_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

● 水体影响

$$C_4 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 \end{bmatrix}$$

● 固体废弃物影响

$$C_5 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 \end{bmatrix}$$

● 人体危害

$$C_6 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \end{bmatrix}$$

■ 确定权重

根据层次分析法的算法可以得到各级指标的权重分别为：

$$\begin{aligned} C_1 &= [0.5, 0.5]^T, \\ C_2 &= [0.31, 0.07, 0.31, 0.31]^T, \\ C_3 &= [0.33, 0.33, 0.33]^T, \\ C_4 &= [0.13, 0.37, 0.13, 0.37]^T, \\ C_5 &= [0.13, 0.13, 0.37, 0.37]^T, \\ C_6 &= [0.13, 0.13, 0.37, 0.37]^T. \end{aligned}$$

■ 评分

然后，将对空调的环境属性的各项指标按照表 37.4-4 的评分标准对四级指标进行打分。打分结果如表 37.4-5 所示。

根据加权平均法，我们可以得到三级指标的分值：能源消耗  $B_1 = 0.5 \times 1 + 0.5 \times 3 = 2$ 。同理，资源消耗、大气影响、水体影响、固体废弃物影响、人体危害的分值分别为： $B_2 = 2.07$ ； $B_3 = 3.3$ ； $B_4 = 5$ ； $B_5 = 2.63$ ； $B_6 = 3.02$ 。

■ 分级评价

为了直观地表示空调绿色度的好坏，先将空调器的经济属性、技术属性和环境属性量化。此处采用如表 37.4-6 所示的 5 级分制。

● 模糊评价矩阵 B 的确定

由上述方法，二级指标环境属性下面的三级的分值均已确定，应用层次分析法，可确定 {能源消耗、资源消耗、大气影响、水体影响、固体废弃物影响、人体危害} 对环境属性的权重，此即为模糊评价矩阵 B。

表 37.4-4 评分标准

指标名称	评分标准	得分
能源消耗量	达到国际先进水平	5分
	达到国家先进水平	3分
	达到国家标准	1分
	此空调只达到了国家标准	
能源利用率	空调的制冷系数（又称能效系数或效能比）大于 3	5分
	空调的制冷系数接近 2.7 （因为 2500~4500W 的空调，效能比系数的国家标准是 2.7）	3分
	空调的制冷系数小于 1.0	1分
	其中制冷系数 = 铭牌制冷量 / 铭牌输入功率，本试验的空调制冷系数是：3500/1350 = 2.59，故能源利用率量打分为 3 分	
资源消耗量	达到国际先进水平	5分
	达到国家先进水平	3分
	达到国家标准	1分
资源利用率	其值超过 80%	5分
	其值接近 50%	3分
	其值低于 30%	1分
重复再用率	重复再用率大于 60%	5分
	重复再用率接近 40%	3分
	重复再用率低于 10%	1分
再生性	再生率大于 60%	5分
	再生率接近 40%	3分
	再生率小于 10%	1分
破坏臭氧层物质	制冷剂的 ODP（臭氧破坏潜能值）小于 0.05	5分
	制冷剂的 ODP 大于 1.0	1分
	其余的看 ODP 值的大小可打分为 2、3、4 分	
温室效应的物质	制冷剂的 GWP（温室效应系数）小于 0.5	5分
	制冷剂的 GWP 值大于 20	1分
	其余的看 GWP 值的大小可打分为 2、3、4 分	
存放性	对于固体废弃物的存放无特殊要求，存放容易	5分
	对于固体废弃物的存放要求有特殊的容积或场地	1分
	其余情况可按其存放要求的高低分别打为 2、3、4 分	
降解性	固体废弃其自然降解年限小于 0.5 年	5分
	固体废弃其自然降解年限大于 5 年	1分
	其余情况可按其降解年限的高低分别打为 2、3、4 分	
噪声	室内机和室外机的噪声均小于 35dB	5分
	室内机噪声超过 45dB 小于 60dB，但室外机噪声小于 35dB	4分
	室内机噪声超过 35dB 小于 45dB，但室外机噪声小于 45dB	3分
	室内机噪声超过 35dB 小于 45dB，室外机噪声小于 45dB 小于 60dB	2分
	室内机噪声超过 45dB 且室外机的噪声超过 60dB	1分

注：对于形成酸雨的物质、有机无毒物、有机有毒物、无机无毒物、无机有毒物、固体排放、存放渗出液、粉尘、辐射、有毒物质这几个指标由于空调产品并不产生这方面的污染物所以均打分为 5 分。

表 37.4-5 环境属性四级指标分值

指 标	分 值
能源消耗量	1
能源利用率	3
资源消耗量	1
资源利用率	3
再用性	2
再生性	3
破坏臭氧层物质	2
形成温室效应的物质	3
形成酸雨的物质	5
有机无毒物	5
有机有毒物	5
无机无毒物	5
无机有毒物	5
存放性	5
降解性	1
固体排放	5
存放渗出液	5
粉尘	5
噪声	3
辐射	5
有毒物质	5

表 37.4-6 评分标准

好	良	一般	差	极差
5分	4分	3分	2分	1分

根据判断矩阵的构建，这三级指标的判断矩阵如下：

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 2 & 2 & 1 \end{bmatrix}$$

由求和法计算可知该矩阵的特征向量为  $S =$

$[0.14, 0.14, 0.14, 0.14, 0.14, 0.30]^T$ ,  $S$  为各三级指标对环境属性的权重。

● 模糊矩阵  $R$  的确定

根据用 0-1 标度法：

□ 对于落于两端点之外的指标，取其隶属度为 0；

□ 对于落于区间  $(A, B)$  的指标，则用下述公式确定隶属度  $X_1, X_2$ 。

$$X_1 = 1 + A - \text{实际值}$$

$$X_2 = 1 - B + \text{实际值}$$

这里的  $A$  和  $B$  是表 4-6 中的各个分值,  $A, B=1, 2, 3, 4, 5$ ,  $X_1$  为属于  $A$  这一级的隶属度,  $X_2$  是属于  $B$  这一级的隶属度。

按照这一方法,即可确定能源消耗的隶属度为  $[0, 0, 0, 1, 0]^T$ ; 同理,资源消耗、大气影响、水体影响、固体废弃物影响、人体危害的隶属度分别为  $[0, 0, 0.07, 0.93, 0]^T, [0, 0.3, 0.7, 0, 0]^T, [1, 0, 0, 0, 0]^T, [0, 0, 0.63, 0.37, 0]^T, [0, 0.02, 0.98, 0, 0]^T$ 。

即

$$R = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.07 & 0.93 & 0 \\ 0 & 0.3 & 0.7 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.63 & 0.37 & 0 \\ 0 & 0.02 & 0.98 & 0 & 0 \end{bmatrix}$$

■ 综合评价

确定环境属性的分值：

$$A = S \times R = [0.140, 0.045, 0.490, 0.332, 0]^T$$

$$\text{Max} \{a_1, a_2, a_3, a_4, a_5\} = 0.49 = a_3$$

由此可以得出结论,此空调的绿色度的环境属性为 3 分。按此得分,其等级属于一般。

## 第5章 绿色包装设计

包装在产品的仓储、运输、运销方面,起着极其重要的作用,同时包装对环境也有着不容忽视的影响。因此,在绿色产品的设计中,也必须对包装给以应有的重视。

### 5.1 产品包装设计的现状

#### 5.1.1 产品包装的基本功能

在现代产品流通领域中,产品的包装,特别是产品的外包装,具有十分重要的地位。因为产品的外包装不啻是产品的一件漂亮的外衣,这犹如人们外在的穿着打扮,以显示其气质、情趣、修养、风度、身价和地位。因此,产品的包装便往往能够吸引消费者的购买欲,提高产品的竞争价位,乃至增强企业的品牌效应和市场竞争能力。这也许是一些企业乐意花费大量投资于产品包装的原因。据报道,全球每年用于产品包装的费用达500亿美元。当然,产品包装的目的并不在于此。产品包装的主要目的,还在于它具有以下重要的功能:

- 确保产品在到达消费者手中时,是处于完好无损的状态。

- 当产品在仓储(保管)和配送(运输)时,应使产品的损坏和报废减小到最小程度。

- 必须保证产品的卫生与安全(特别是对于食品与药品一类的商品,由于顾客害怕污染与不洁,有损健康,因此这一点变得尤为重要)。

- 应便于产品的运输、搬移、堆垛和存放。

- 应便于消费者迅速识别商品(这一点对于商品种类繁多的自助式超级市场,则尤为重要,因为这时同类商品很多,而且又同时集中摆放在相邻货架上,所以要能让消费者一眼便识别出所欲购买的商品,则醒目的包装就显得特别重要)。

- 包装也是生产厂家与消费者进行信息交流沟通的重要手段。通常有关产品的属性、使用和保养方法的说明,以及与厂家联系的地址、电话等信息,都可印刷在产品的包装上。对于一些小商品,这一功能更为重要。因为许多消费者正是通过阅读包装上的有关信息介绍,才引起他们的购买欲。

- 同样,包装还为仓储与运输提供所需要的,诸如如何放置、吊装、防潮、防震、防晒等信息。这些信息可以用图示标记或文字印刷在外包装(包装容器)的醒目处,以引起搬运人员的注意。

目前在产品包装设计中,还存在着所谓的“过度包装”的问题。这主要是生产商旨在应对激烈的市场竞争而产生的一个问题。为了吸引顾客购买自己的产品,不少生产商不惜工本在产品包装上做足文章。这时,包装设计所考虑的问题,已经不是如上所述的那些包装的基本功能,而是过分强调了包装的美学观点。由于这种多少有些本末倒置的设计思想,因此陡增了包装的成本;甚至有的产品的包装成本,远远高于产品本身的成本。这种情况多见于轻工产品,特别是美容化妆品等。

实际上,产品包装设计的基本原则,应该是经济、实用、美观。首先应该考虑包装的经济性,以便使所设计的产品包装尽可能地经济,从而不会因包装过分而增加产品的成本。其次则要求达到包装的实用性,使所设计的产品包装,能达到产品本身对包装提出的基本功能要求,并力戒那种既不实用、又陡增成本的过度包装。最后才是研究包装的美观性,也即在满足产品包装的经济性和实用性的前提下,适当考虑包装的美观问题。必须指出的是,产品过度包装所增加的费用,生产商是不会承担的,而最终还是转嫁给了消费者。

#### 5.1.2 产品包装引起的环境问题

正如以上所述,包装只是产品的一层“外衣”,它除了在商品流通领域的运输、仓储、配送过程中,使商品免受损伤外,严格地说,便无其他功能。因此,当商品一旦到达消费者手中后,包装便完成了它的主要使命而寿终正寝。这是因为消费者购买的是商品的使用价值,而非商品的包装,所以拆除包装后,包装便一无用处。这时的包装便成为一种废弃物,一般被当作垃圾处理。

根据欧洲和北美大多数国家的统计数据显示,城市固体废弃物(MSW: Municipal Solid Waste)中,有30%来自被废弃的商品包装材料。由此可见,包装材料对增加环境负担的影响是很大的。

我国每年固体废弃物的排放量约为6亿t,其中40%是可以回收利用的,而其余60%则属于无法处置或难于处置的。在这些固体废弃物中,包装废弃物也同样占了约1/3。这对于我国的生态环境构成了极大的危害。如今,不少城市已对如何处置这类与日俱增的废弃物,感到十分困惑和棘手。可以填埋废弃物的场地已日益紧缺。为此,有的城市采用垃圾焚烧的方法,不惜投资兴建垃圾焚烧厂。个别地区由于废弃物处置不当,

甚至还引发了莫名其妙的地方病,严重损害了人民群众的身体健康。

塑料的诞生,对于包装行业而言,不啻掀起一起包装材料的革命。在过去的 30 多年中,塑料在包装材料业界,可以说是独领风骚。在人们的日常生活中,无处不见塑料包装,也无处不用塑料包装。据统计,目前全世界塑料年产量已逾 1 亿 t 以上。用作包装材料的塑料约占其中的 1/3。塑料包装物被废弃后,往往以下述两种方式出现:

- 作为城市固体废弃物,被纳入规范的城市垃圾处置系统;

- 被人们随意抛弃,而形成严重的“白色污染”。

由于塑料不能自行降解,而在目前所有的包装材料中,塑料这种包装材料的回收率又极低;加之,塑料无法进行焚烧处置(因为塑料焚烧时,将产生破坏臭氧

层的化学物质),因此上述两种方式,最终都将给人们的生存环境带来污染。

总之,产品包装所产生的环境问题,必须引起人们的足够重视。

## 5.2 绿色包装设计

传统的包装设计,并不从包装的全生命周期去考虑。它只是为包装而作包装设计。至于完成包装使命后,对于包装应如何处置,则似乎已不是它所应考虑的一份内事了。绿色包装设计却与此相反,它不仅要考虑包装必须满足的基本功能,同时还必须考虑包装完成使命后的回收和再利用问题,以杜绝和防止包装对环境造成的负面影响。因此,绿色包装设计与传统包装设计是有区别的。图 37.5-1 表示了绿色包装设计的流程图。

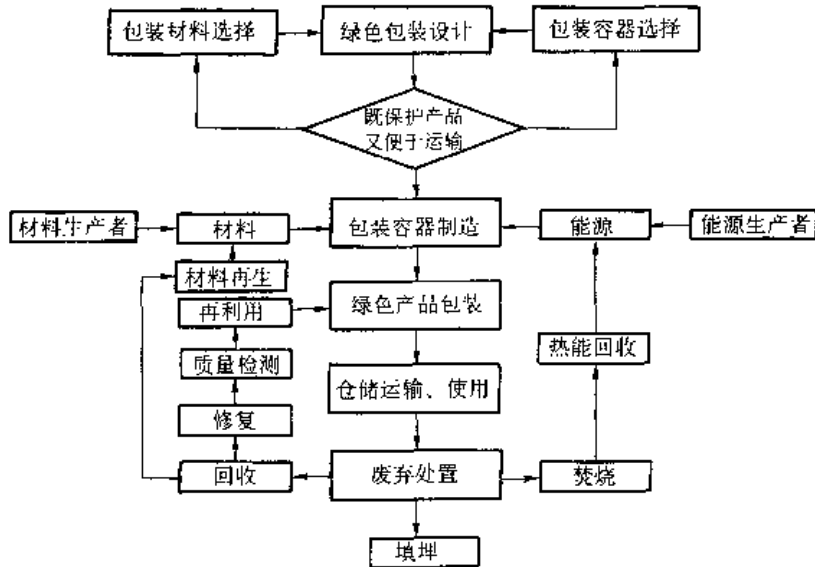


图 37.5-1 绿色包装设计的流程图

由图中可知,流程图的中央部分,是绿色包装设计按其生命周期的主要工作流程。它包括:包装设计→包装制造→包装使用→废弃处置等 4 个主要阶段。产品的包装被废弃后,有 3 条出路,即回收、填埋和焚烧。当然优先考虑的,是回收被废弃的包装,以便使包装能多次重复循环利用;当原包装无法重复循环利用时,便只能作为材料回收后再生利用。废弃包装的回收路线,可从图中左侧部分看出。当废弃的包装既不能循环使用,又不能作二次材料利用时,此时需视情况处置:若可以焚烧,则可通过焚烧而回收热能;若无法焚烧,则只能用填埋来处置。废弃包装的焚烧路线,即如图中右侧所示。

在图 37.5-1 中的最上面一行,实际上还提示了绿色包装设计所包含的主要内容,即包装容器的选择和包装材料的选择。因为在设计产品的包装时,首先要确

定包装容器的类型和制造包装容器的材料。判断包装设计是否合理,主要是看它能否满足保护产品和便于运输的要求。如果不能满足,则需重新选择容器类型和容器材料,重新进行包装设计,直至满足上述要求为止。下面便介绍绿色包装设计的具体内容和步骤。

### 5.2.1 绿色包装设计的内容和步骤

#### ■ 绿色包装设计的依据

首先应了解被包装产品的基本属性和参数。因为产品的这些属性和参数,将直接影响包装设计。有关产品的这些属性和参数,主要有:

- 产品特征参数:产品的重量、外形及其外形尺寸。

- 产品包装特性:是否需要耐寒、耐高温、耐腐蚀、耐压、抗震、抗湿、抗潮、防水、防晒等。这些特



性显然是和产品仓储与运输的环境有密切关系。

- 产品运输方式：采用单一的公路（汽车）、铁路（火车）、航空（飞机）、航运（船舶）方式，或联运的方式。

- 装卸搬运方式：采用人工、吊车、叉车、或传送带。

通过以上所列举的产品参数和属性，可以明显看出，它们对于包装容器的设计和包装材料的选择，有着

至关重要的影响。因此，在着手设计产品包装之前，必须充分掌握好产品的上述参数和属性。

#### ■ 绿色包装方案设计

绿色包装设计的内容包括包装材料选择、包装容器造型和结构设计、防护包装（又称：内包装）设计、包装废弃物回收处置方法确定以及包装成本核算等。如图 37.5-2 所示，即形象地说明绿色包装设计的具体内容和步骤。

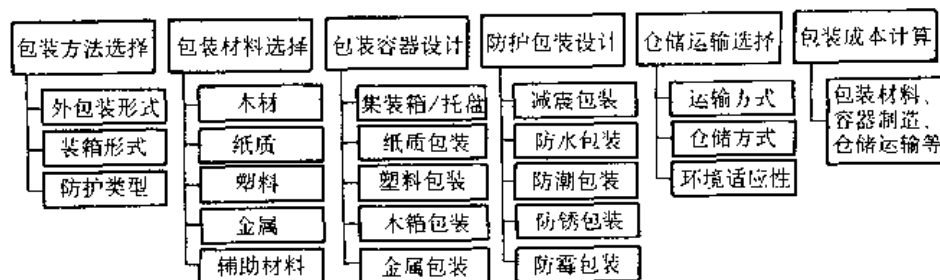


图 37.5-2 绿色包装设计主要内容和步骤

这里需要说明的是，一般包装有外包装和内包装之分。外包装属于包装容器的范畴；而内包装则是在包装容器内部另行设置的包装，即属于此处所谓的防护包装的范畴。内包装的功能，主要是使产品能在包装容器内正确可靠地定位而起到填充的作用，同时按产品包装特性要求，还应起到承受冲击振动的缓冲作用和防水、防潮、防锈等的防护作用。

### 5.2.2 绿色包装的设计准则

#### ■ 绿色包装材料选择

包装材料的选择，是一个使人模棱两可，甚至有时是不知所措的问题。因为有的包装材料，虽然在其生产时需要消耗较多的能源，或者需要消耗不可再生的资源，但是却具有较长的使用寿命；有的包装材料容易降解处理且无害；有的包装材料在焚烧处理时，可以回收其热能，然而却会污染大气。由此可见，迄今为止，尚未找到一种对环境完全友好的包装材料。尽管如此，这并不妨碍人们总结出一些合理选择包装材料的基本准则。

- 尽可能选择无毒性的包装材料 这样可使这种包装材料在其生产、使用和最终处置时，不会对环境对人体健康产生有害影响。

- 尽可能选用资源丰富的包装材料 资源丰富，意味着生产这种包装材料的原料充足，也即易于获得这种包装材料，因而可以降低产品的包装成本。

- 尽可能选用可重用或可再生的包装材料 由于包装材料是可重用或可再生的，因此包装材料可以多次重复循环利用。这样有利于降低产品包装成本。目前可重用和可再生的包装材料有纸、纸板、玻璃、金属

和某些塑料。

- 尽可能减少包装材料的种类 在设计产品的包装时，尽量选用单一的包装材料。这样可以便于日后经济而合理地回收和最终处置。否则，将增加回收时对不同包装材料的分拣费用。

如图 37.5-3 所示，即为 3 类常用的绿色包装材料：可回收再利用的，可自然风化回归的和可焚烧回收热能的。这实际上也是包装材料的三种回收处置方式。三者的排列顺序，也即选择包装材料时的优先顺序。由此可知，应该优先选用可回收再利用的材料，其次是选用可自然风化回归的材料，最后选用可焚烧回收热能的材料。

#### ■ 包装容器的造型与结构设计

合理设计产品的包装容器结构，是从根本上节省资源、减少污染和降低产品包装成本的有效途径。这个设计环节也最能体现包装设计人员的聪明才智和创新能力。包装容器的类型，如图 37.5-2 中所示。设计绿色包装容器的准则有：

- 应使包装结构小型化和轻量化 包装结构的小型化和轻量化，将显著节约包装材料，降低包装成本，而且也便于搬移和运输，从而降低储运成本。

- 应使包装结构便于回收重用 这是产品包装设计的发展方向。正如前面 1.3.1.2 节中列举的 3M 公司的例子所示，3M 公司由于采用了新型高密度聚乙烯包装箱，使得这种包装箱可以循环使用多达 49 次。这里既有包装材料的选用问题，也有包装容器本身结构设计的问题。只有正确地综合解决这两方面的问题，才能使包装容器得到长效的回收重用。惟其如此，才能最大限度地节约资源，降低企业的经营成本。

● 尽量使包装结构简单化和简装化 目前不少包装设计过于花哨,一味片面地追求美学上的视觉效果,以吸引顾客的购买欲望,从而使包装费用陡增,甚至有时还造成资源的极大浪费。

值得注意的是,据美国的一份资料报道,预计今后40%的商品与服务,将由政府部门和其他社会集团集中采购,而非私人或个体消费者所购买。因此,对于这类社会集团集中采购的商品,其包装即可大大简化。我国政府部门以及一些社会集团,也已开始通过招标方式集中采购。这种集中采购的方式,不仅能降低集团的采购成本,杜绝采购中的不正之风;同时也给生产厂家简化包装提供条件,从而降低包装费用,降低产品总的成本。这是一种一举两得的双赢做法,值得大力倡导。

● 在产品包装方法选用上,应尽量按照下列优先顺序

- 无包装
- 最简单包装
- 可返回/可重用/可再灌装的包装
- 可回收利用的包装

按照上述优先顺序来选用包装方法,即自上而下地优先选用无包装,最后才选用可回收利用的包装。这一选用产品包装方法的优先顺序政策,是美国东北部州长会议,为了更好地解决产品包装废弃后对环境造成巨大的危害而制订的。他们所制订的这一政策,我们以为值得借鉴,因此将其收录在此,作为产品绿色包装容器造型与结构设计的准则之一。

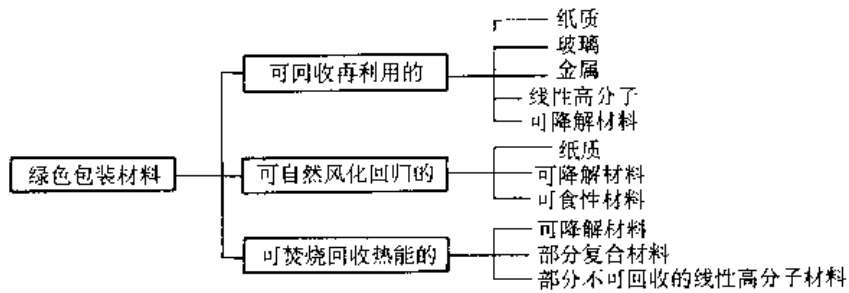


图 37.5-3 常用绿色包装材料及其回收方式

# 附录

## 附录 1 废弃产品最终处置时受到限制的材料

下列表格中列出的材料,在最终处置时,应倍加注意、分外小心。因此,在产品的设计选用材料时,也应慎用或少用。

### 金属材料

中文名称	英文名称
锑	Antimony
砷(及其化合物)	Arsenic (and its compounds)
钡	Barium
铍(及其合金)	Beryllium (and its alloys)
镉	Cadmium
铬化物 (特别是六价铬)	Chromium compounds (especially hexavalent)
钴	Cobalt
铜	Copper
铅(及其化合物)	Lead (and its compounds)
锂	Lithium
镁	Magnesium
汞	Mercury
钼	Molybdenum
镍	Nickel
硒	Selenium
银	Silver
碲	Tellurium (and its compounds)
铊	Thallium
钒	Vanadium
锌	Zinc

### 无机材料

中文名称	英文名称
石棉	Asbestos
二硫化碳	Carbon disulfide
氰化物	Cyanides
氟化盐	Fluoride salts
氟硼酸盐	Fluoroborates

### 有机材料

中文名称	英文名称
2,3,4-涕丙酸	2,4,5TP (Silves)
丙酮	Acetone
苯	Benzene
四氯化碳	Carbon tetrachloride
强力杀虫剂	Chlordane
氯苯	Chlorobenzene
氯仿	Chloroform

(续)

中文名称	英文名称
甲醛	Cresols
甲酚酸	Cresylic acid
环己酮	Cyclohexanone
二氯代苯	1/4 Dichlorobenzene
二氯乙烷	1,2-Dichloroethane
二氯乙烯	1/1-Dichloroethylene
二硝基甲苯	Dinitrotoluene
异狄氏剂	Endrin
2-乙氧基乙醇	2 Ethoxyethanol
乙酸乙酯	Ethyl acetate
苯乙烷	Ethyl benzene
乙醚	Ethyl ether
七氯	Heptachlor
六氯苯	Hexachlorobenzene
六氯丁二烯	Hexachlorobutadiene
六氯乙烷	Hexachloroethane
异丁醇	Isobutanol
林丹	Lindane
甲醇	Methanol
二丁酮	Methyl ethyl ketone (2 Butanone)
甲基乙丁醇	Methyl isobutyl ketone (4-Methyl-20pentanone)
甲基氯	Methylene chloride
甲氧氯	Methoxychlor
硝基苯	Nitrobenzene
二硝基丙烷	2-Nitropropane
五氯苯酚	Pentachlorophenol
吡啶	Pyridine
聚四氟乙烯	Teflon (if heated or machined)
四氯乙烯	Tetrachloroethylene
甲苯	Toluene
毒杀酚	Toxaphene
三氯乙烯	Trichloroethylene
三氯(苯)酚	2,4,5-Trichlorophenol
三氯苯酚	2,4,6-Trichlorophenol
氯化乙烯	Vinyl chloride
二甲苯	Xylene

## 附录 2 国外用于 LCA(生命周期分析)的工具软件一览表

列出这份软件一览表的目的是,是便于供读者今后如有需要,可与这些机构进行联系。

LCA 工具名称	工具开发者	国别	类型	影响评价	用于复杂产品	数据可靠性
Pre LCA Tool	Battelle/Digital	美国	半量化	是	是	不
Matrix Approach	AY&Y	美国	半量化	是	是	不
CUMPAN	Daimler Benz	德国	量化	是	是	是
EPS-Tool	Swedish Envir. Reas. Inst. (IVL)	瑞典	量化	是	是	?
EDIP-Tool	Inst. for Prod't. Devel't	丹麦	量化	是	是	是
IDEA	VTT(SF)	南非	量化	是	是	是
PIA	Inst. Voor Toegepaste	荷兰	量化	是	?	是
SimaPro	Pre Consulting	荷兰	量化	是	是	是
EcoAssessor	Pira International	英国	量化	?	?	是
DEMS	PIRA International	英国	量化	是	是	是
Oeko-base	Migros	瑞士	量化	是	不	是
LMS Eco. Inv. Tool	LMS Umwelt System	瑞典	量化	?	?	是
Eco-Pro	EMPA	瑞士	量化	是	?	?
GaBi	Inst. for Polymer Test & Scien.	德国	量化	是	是	是
TEAM	Eco-bian	法国	量化	是	是	是
Lims	Chem. System	美国	量化	?	?	是
LCA Invent. Tool	Chalmers Industriteknik	瑞典	量化	不	是	是
Houstead Model	Houstead Consulting	英国	量化	不	是	是

附录 3 国外开发的用于产品生命周期分析(LCA)的数据库(环境影响因素清单分析用的数据)

数据库名称	重点	来源	可用性
ETH ESU 96	主要侧重于电力生成和相关过程,如传输、加工、废弃物处理。其中包括了 500 种最重要的过程。此外还提供了扩展文档	ETH-ESU	包括在 SimaPro 标准数据库中(有商用版本和教学版本)
BUWAL 250	重点在包装材料(塑料、纸板、纸、玻璃、镀锌钢、铝)、能源、运输、废弃物处理	BUWAL 250 (第 2 版)(BUWAL 是瑞士内政部环境局)	包括在 SimaPro 标准数据库中(有商用版本和教学版本)
INDUSTRY DATA	由工业部门提供的环境影响因素清单,大多数属于企业内部的数据	APME,以及来自各方面	包括在 SimaPro 标准数据库中(有商用版本和教学版本)
IDEMAT 2001	重点在工程材料(金属、合金、塑料、木材)、能源、运输	由荷兰 Delft 工业大学工业设计工程系指导下采集数据	包括在 SimaPro 标准数据库中(有商用版本和教学版本)
FRANKLIN US LCI	用于能源、运输、钢、塑料、加工工艺的北美环境影响因素清单数据	数据由美国 Franklin/公司采集	包括在 SimaPro 商用版本数据库中。教学版本需单独购买
Data archive	包括材料、能源、运输、加工工艺、废弃物处理	各种早期公开发表的数据源(PWMI, BUWAL132, ETH, SPIN, Chalmers, Kemna)	包括在 SimaPro 商用版本数据库中。教学版本需单独购买
Dutch Concrete Database and scripts	重点是荷兰具体生产和使用中所有方面的有关数据。可结合说明书使用。数据与说明书均为荷兰文	荷兰 Betonplatform	拥有服务合同的荷兰用户是免费的。其他拥有服务合同的客户,按申请可得,但无语言支持。这一数据库受到复制和出口的保护
IVAM 3.0	包括材料、运输、能源和废弃物处理。大部分是荷兰的数据	数据是由荷兰阿姆斯特丹 IVAM 环境研究所收集的。其中包括 ETH/BUWAL 和 Pre 的数据	需向 IVAM 环境研究所购买
FEFCO	主要关于欧洲的瓦楞纸板生产数据,部分是根据上述 BUWAL 250。含有瓦楞纸板生产和生命周期建模的说明书	FEFCO,欧洲瓦楞纸板制造商协会(总部在法国)	只凭 FEFCO 许可证随 SimaPro 一起供应。可免费获得相随报告

#### 附录4 国外从事生态设计(Ecodesign)或环境意识设计与制造(Environmentally Conscious Design and Manufacturing)的机构名录

列出这份机构名录的目的,是便于供读者今后如有需要,可与这些机构进行联系。

序号	机构名称
1	Redesign
2	Codesign journal(Green design section)
3	Sustainable Development Network Directory
4	02 (Flash 4 needed)
5	Indes
6	BQUEST (Working towards sustainable urban development)
7	Eco-efficiency and Cleaner Production Australia
8	Design for sustainability program of Technical University Delft
9	Consortium on Green Design and Manufacturing USA
10	National Centre for Environment Design at RMIT Australia
11	OECD Environmental Program
12	Centre for Sustainable Design UK
13	EcoDesign Canada
14	EcoCycle Canada
15	Design for Environment US EPA (EPA 是指美国环境保护署)
16	Sustainable source from Greenbuidr USA
17	Mapping the journey (Case studies on sustainable design by Ekos International, Canada)
18	Ecodesign (Austrian Ecodesign Information point)
19	Greenerbydesign (Examples of green design, quiz, forum and more)
20	EnvironmentalExperts (Environmental Business Portal)
21	Sustainable development library (Comprehensive list on sustainable development)
22	NatLogic (Extensive environmental link list)
23	Enviroreporting (on environmental reports world wide)
21	The global Ecolabelling Network

#### 附录5 国外从事生命周期评估(LCA)的机构名录

列出这份机构名录的目的,是便于供读者今后如有需要,可与这些机构进行联系。

序号	机构名称
1	Life Cycle Links (by Thomas Gloria, USA)
2	LCA hotlist (by Gabor Doka, Switzerland)
3	LCAccess (by US EPA)
4	SPOLD (Society for Promotion of Life-cycle Assessment Development)
5	SPINE (Swedish Life Cycle inventory data for sale, and more)
6	LCANET (European Network for Strategic Life Cycle Assessment Research & Development)
7	CHAINET (European network on chain analysis for environmental decision support)
8	LCA in Japan (OVerview of LCA activities in Japan)
9	APME (Ecoprofiles of plastics-registration required)
10	Sylvatics (LCA software tools)
11	How products impact (illustrates the relationship between products and natural ayatems with a simple example. Nice educational resource.)
12	Environmental resources
13	Environmental Experts (Environmental business Portal)
14	Sustainable development library (comprehensive list on sustainable development)
15	NatLogic (extensive environmental link list)
16	Enviroreporting (on environmental reports world wide)
17	The global Ecolabelling Network
18	Envirotech (Strategies with innovative environmental technology)
19	Mineral Resources Forum (Minerals, netals and sustanibale development)

#### 附录6 有关绿色产品设计的相关网站

以下列出的资料是,介绍在绿色产品设计领域中的一些非赢利的公开组织,或与私营组织一起工作的学术机关的组织概况、产品与服务、网站及联系方式。以供读者有需要检索时参考。凡赢利的机构此处均不予列出。

##### 1. Business for Social Responsibility Education Fund (BSR)

###### ■ 组织概况

BSR(为社会责任而经营)是总部设在美国的一个全球性的机构,它为旨在保持商业成功的公司,继续寻求尊重人、尊重社区和尊重环境的途径。通过BSR的会员制,各个公司会员已经得到有关企业社会责任的

实用信息、研究、教育和培训计划以及技术帮助和技术咨询。BSR 拥有 1400 家以上的公司会员或分支机构,代表着 600 万工作人员和 1.5 万亿美元的收入。BSR 的附属组织:BSR 教育基金会是一个非赢利的慈善组织。它为更广泛的商务团体和有关一般公开的负责责任的商务实践的研究与教育计划服务。BSR 的经营与教育计划(BATE)是一个教育机构。它帮助许多公司把对环境的考虑结合在其战略经营决策中。它还帮助创建了更多的在环境上是可持续的商业系统。BATE 启动了包括全球气候变暖、绿色产品设计、绿色建筑、可持续经营实践等相关项目。

#### ■ 产品与服务

BATE 与许多公司在不同行业内一起工作,以便把面向环境设计(DfE)的最好实践经验,引入它们的产品开发过程中;从而有助于公司在它们的产品设计中消除环境影响或资源利用率不高的弊病。BATE 的计划提供了下列服务:

- 信息服务 编印出版各种报告、学习资料和检验领导层面在各个领域实践结果的工业(行业)标准。

- 各种讨论会和讲习会 组织引导有关各种能源和气候变暖等话题的讨论会和讲习会,从而把公司和股东们聚集在一起分享经验、资源和知识。

通过 BSR 教育基金会的计划,也为广泛的商务团体、独立机构以及其他单位,提供基本的信息的援助。对于非会员的公司,在付费的基础上,也可以得到比该计划更深入的产品与服务。

#### ■ 网站

BSR 的综合网站含有有关会员、产品、服务、计划、会员公司及其网站的链接等信息。网站还提供了 BSR 的全球经营责任资源中心的在线数据库,这一数据库覆盖了 100 个以上的企业社会责任问题范围。中心的数据库包含了一些公司领导的实践结果、公司政策样板、外部标准、在环保领域中的最新进展以及一份经过挑选的专家名单等有关信息。这一网站还包括了涉及环境方面的话题如气候变暖、能源效率、绿色建筑、绿色动力/可再生能源、绿色产品设计、资源生产率、供应商管理、可持续林业、废弃物减量化、水资源保守/效率以及水质等。

#### ■ 联系方式

Business for Social Responsibility  
Business and the Environment Program  
609 Mission St., 2<sup>nd</sup> Floor  
San Francisco, CA 94105  
United States  
Tel: 415. 537. 0890

Fax: 415. 537. 0889

Email: [membership@bsr.org](mailto:membership@bsr.org)

## 2. Center for Design (CFD) at Royal Melbourne Institute of Technology (RMIT)

#### ■ 组织概况

皇家墨尔本工学院的设计中心位于澳大利亚。该中心在促进、支持、沟通和提供通向更高环境质量及可持续性水准的途径方面,起着至关重要的作用。该中心已经是,而且还将继续是一个为创造性的、富有成效的、并在商业上实用的生态可持续成长的各种方法的实验室。这一设计中心以澳大利亚最大技术大学之一的皇家墨尔本工学院构造环境系为基础。中心的活动包括:

- 开发和展示旨在改善产品与服务的环境品质的新的设计方法、工具和工艺方法;

- 与工业界合作设计可持续的产品与服务,并开发更富战略意义的环境指导方针;

- 向政府部门提出通过设计和革新减少环境影响的政策和计划,并且为未来可持续发展而探索新的设计理念和方案;

- 维护一个在生态设计(EcoDesign)、生命周期评价(LCA)、产品管理方面研究与信息交换的国家与国际的网络。

#### ■ 产品与服务

该中心是澳大利亚唯一一个拥有有关各种前沿的工业环境管理工具、方法和过程的积极而完整规划的研究机构。这些工具、方法和过程涉及:

- EcoDesign(生态设计),制造类产品和服务的可持续产品开发;

- 产品、材料、服务和建筑等生命周期的环境评价;

- 产品管理和产品生命周期管理。

#### ■ 网站

网站包含大量有关产品生态设计方面的信息,其中包括许多设计项目的案例和其他大量绿色产品设计资源的链接。除 CFD 主要网站外,下列网站也提供有关绿色产品设计的信息:

- <http://ecospecifier.rmit.edu.au>

- <http://onsite.rmit.edu.au>

#### ■ 联系方式

Center for Design at RMIT  
Royal Melbourne Institute of Technology  
GPO Box 2476V  
Melbourne Vic 3001

**Australia****Phone:** +61(0)3 9925 3484**Fax:** +61(0) 3 9639 3412**Email:** [john.gertsakis@rmit.edu.au](mailto:john.gertsakis@rmit.edu.au)**3. Center for Sustainable Design (CfSD)****■ 组织概况**

可持续设计中心设在英国,而隶属于萨里艺术与  
设计研究所及萨里与汉普郡环境事务协会。该中心推  
动有关生态设计以及在产品和服务开发设计中的环  
境、经济、伦理与社会关注方面的研究和讨论。中心还  
扮演着信息交换者的角色,并且成为可持续产品和服务  
革新思想的焦点。

**■ 产品和服务**

CfSD 的产品和服务包括涉及绿色产品设计的讲  
习会、讨论会以及其他类型的培训与教育、研究、会议、  
咨询和出版物等。

**■ 网站**

它提供有关各种项目的信息和各种在线的出版物  
与研究报告。

**■ 联系方式****The Center for Sustainable Design****Faculty of Design****The Suerey Institute of Art & Design****Fslkner Road****Farnham, Surrey GU9 7DS****United Kingdom****Phone:** +44(0)1252. 892772**Fax:** +44(0)1252. 892747**Email:** [cfsd@surrart.ac.uk](mailto:cfsd@surrart.ac.uk)**4. EcoMarket International****■ 组织概况**

国际生态市场(EcoMarket International)是一个  
在线的参考资源。它已发展成为可供任何对设计、认  
证、制造或采购绿色产品感兴趣的人使用的公共服务  
机构。

**■ 产品和服务**

Ecomarket 是一个与提供各种绿色产品与服务企  
业链接的在线可搜索数据库。虽然这个数据库重点在  
欧洲,但是它也提供了与全世界许多企业的链接。

**■ 网站**

此网站提供了几百个其他网站的链接。数据库中  
包括了以下数据类别:生态设计/革新、工业/技术、建  
筑/家具、服装/纺织品、家用电器/化妆品等等。

**■ 联系方式**

EcoMarket 没有邮件地址、没有电话或传真号码。  
只能通过它的因特网地址(URL)取得联系

**<http://www.ecomarket.net>****5. Green Design Initiative****■ 组织概况**

卡内基梅隆大学的绿色设计创始者(Green  
Design Initiative,以下称其为绿色设计研究所),倡导  
了环境意识的工程、产品与工艺设计、制造以及建筑。  
这个研究所与许多工业企业、基金会和政府部门形成  
合作伙伴关系,以便发展联合的研究和教育计划,在支  
持可持续发展的经济发展的同时,来改善环境质量。绿色  
设计研究所把注意力集中在通过回收稀有资源、采用更  
少的原材料而创造更好的产品,来开发实用的防止污  
染和降低成本的技术。

**■ 产品和服务**

绿色设计研究所和工业界合作开发产品,并共享  
案例及最佳实践信息。这项计划也开发了绿色设计工  
具,且出版了许多学习资料、报告和论文。这些工具和  
出版物均可在线订购。

**■ 网站**

绿色设计研究所的网站提供了有关其各种研究项  
目和工业界合作计划、有关绿色设计出版物的明细表  
以及与其他相关因特网网站的链接等信息。

**■ 联系方式****Green Design Initiative****Carnegie Mellon University****5000 Forbes Avenue****Pittsburgh, PA 15213-3890****United States****Phone:** 412. 268. 2299**[http://www.ce.cmu.edu/Green\\_Design](http://www.ce.cmu.edu/Green_Design)****6. Green Seal****■ 组织概况**

绿色印记(Green Seal)是一个独立的非盈利组  
织。它是专门通过倡导制造与销售对环境负责任的消  
费品而保护环境的。它设立了各种环境标准和一种认  
同为绿色印记的奖项,奖给那些比同类产品对环境伤  
害要小的产品。绿色印记的环境伙伴计划,给关注其采  
购环境影响的商家们,提供绿色的购买援助。

**■ 产品和服务**

● 为大约 50 个门类以上的对环境更好的产品  
提供推荐;其中包括打印用纸和复印用纸、油漆、节能  
照明、再次提纯机油和节水装置等。

● 出版“办公室的绿色采购指南”。它包括关于

如何实施一项采购政策以及推荐对环境更好的产品等方面的信息。

● 出版“挑选绿色”的报告。这一报告中列出了对环境负责的产品清单,并且表示其品牌和货源,其中包括针对复印纸、打印纸和不用树生产的纸的专题报告等。

#### ■ 网站

绿色印记网站含有关于各种绿色印记计划和一个供参加绿色印记伙伴计划的在线注册登录的格式等背景信息。该网站还录有绿色印记各种标准的摘要。

#### ■ 授奖

● 授予绿色印记奖的产品,可将其用于包装、广告和促销上。

● 参与绿色印记伙伴计划的公司,可以将绿色印记伙伴标识图案用于公司的信封、信笺和广告上,以分担它们与顾客、雇员和公众的环境承诺。

#### ■ 联系方式

##### Green Seal

1001 Connecticut Avenue, NW, Suite 827

Washington, DC 20036-5525

United States

Phone: 202. 872. 6400

Fax: 202. 872. 4324

Email: [green seal@green seal.org](mailto:green seal@green seal.org)

## 7. INFORM, Inc.

#### ■ 组织概况

英佛尔姆(INFORM)公司是一个检验企业经营实践结果及对环境和对人类健康影响的非盈利的环境研究和教育组织。它的目标是辨识能保证对环境可持续的经济成长的经营之道。INFORM 公司将其研究集中在三个领域:化学制品的种种危害,资源的保护与运输,以及清洁能源。

#### ■ 产品和服务

INFORM 公司的主要服务是研究、报告和沟通。它用各种书籍、专门报告、业务简讯、论文,还有业务通讯季刊、INFORM 报告等发表其研究。

#### ■ 网站

INFORM 公司的网站除了介绍有关该公司深层次的信息外,还报道了该公司各种主要计划范围内的当前工作。公司的出版物可以在线订购。有些报告、学习资料和论文可在网上获得。

#### ■ 联系方式

INFORM, Inc

120 Wall St.

New York, NY 10005-4001

United States

Phone: 212. 361. 2400

Fax: 212. 361. 2412

<http://www.informinc.org>

## 8. Massachusetts Institute of Technology (MIT), Technology, Business and Environment (TBE) Program

#### ■ 组织概况

建立麻省理工学院(MIT)的技术、经营与环境(TBE)计划,是为了帮助各个公司满足达到卓越环境和经营成功的双重挑战。该项计划的主要活动包括:①分析与评价不同方案掌握和控制有关各种自愿协议、公司合伙关系、报道要求和谈判的规章制度;②检验各种富有创新的环境管理实践如何在一个行业中普及;③研究紧密耦合的工业综合企业、闭环生产系统和公共政策如何进一步应用和实施工业生态学原理。

#### ■ 产品和服务

TBE 计划为数字设备公司(DEC)开发了一种名为面向环境设计入门(DfE Primer)的软件。这项计划还:

● 提供检索有关环境管理与政策方面创新途径的前沿研究,并指出在改善环境与财政绩效中何者有效及何者无效;

● 组织各种诸如生命周期评估和产品回收等研究领域的会议;

● 通过由来自企业界和学术界代表参与的、关于经营与环境的 MIT 座谈会的各种定期会议,来培育合作网络;

● 主持对环境更好的产品与服务的展廊,由此建立了一条展示将近 100 种来自不同行业的产品与服务的在线数据库。这一数据库旨在提供易于检索有关众多最佳实践和创新产品的信息,以便大步推进可持续发展。

#### ■ 网站

DfE Primer 可以在 TBE 网站的出版物区获得。对经营感兴趣的其他学习资料,也可在项目与出版物区得到。对环境友好的商品与服务的展廊则位于网页 <http://tbe.mit.edu/gallery/>。

#### ■ 联系方式

MIT-TBE Program

E40-251

Cambridge, MA 02139-1309

United States

Phone: 617. 253. 3586

Fax: 617. 253. 7140



## 9. The Natural Step

### ■ 组织概况

自然步伐(The Natural Step)是一个非赢利的国际组织。它专注于营造一个对生态和对环境可持续的社会。自然步伐提供了一个以科学为基础的框架,作为供企业、社区、学术界、政府机构以及个人重新设计其活动,使之更加可持续发展的指导方针。这一框架的核心,是由四个系统状态构成的集合,用以描述所有环境问题及其解决方法的科学基础。自然步伐采用系统状态帮助公司以及其他机构,保持其运作与可持续性原则的一致。

### ■ 产品和服务

在美国,自然步伐通过以下方式,帮助各种企业和公司与可持续性相融合:

- 各种活动 举办遍及全国的各种研讨会、执行官简报会、年会、专题讨论会、宣讲会等;

- 培训 指导公司内部的培训以及为一群公司的、政府的和非赢利的领导层的培训,以推动可持续性的四个系统状态和如何能用它创建可持续的企业;

- 信息服务 生产多种工具以帮助公司和组织教育职工、供应商、顾客和其他人群有关可持续性原理。

### ■ 网站

自然步伐网站的作用是,介绍该组织的科研、历史和区域性活动,以及可获得的商品、有关该组织的最新新闻报道和即将举行的各种活动日程表。

### ■ 联系方式

The Natural Step

PO Box 29372

San Francisco, CA 94129

United States

Phone: 415. 561. 3344

Fax: 415. 561. 3345

E-mail: [tns@naturalstep.org](mailto:tns@naturalstep.org)

## 10. 02 Network

### ■ 组织概况

02网络(02 Network)是一个非赢利组织。它弥补了设计专业人员的一个国际网络。02支持介入开发对环境可持续的产品与服务的设计人员和企业家之间的信息与知识交换。

### ■ 产品和服务

网络提供与生态设计人员的链接,以及提供有关生态设计项目、工具和思路方面的各种信息。02也资助生态设计方面的活动。

### ■ 网站

此网站提供与绿色设计专家、工具、思想、书籍、案例以及国际生态设计活动日程表的链接。

### ■ 联系方式

The Neaderlands Design Institute

Keizersgracht 609

1017 DS Amsterdam

The Neaderlands

Phone: +31(0)20. 551. 6500.

Fax: +31(0)20. 6201013

## 11. U. S. Environmental Protection Agency (EPA), Design for the Environment Program

### ■ 组织概况

美国环境保护署(EPA)面向环境设计(DfE)计划,是帮助各个企业把对环境的考虑结合到产品、工艺过程以及技术和管理系统的设计与再设计中去。1992年由EPA下属的防止污染与中毒办公室(OPPT: Office of pollution Prevention and Toxics),通过DfE计划而与工业界、大学、研究机构、公共利益集团以及其他政府部门结成自愿的合伙关系。自此以后,DfE的伙伴计划便存在于许多工业部门,其中包括印刷、电子、服装与纺织、养护、工业洗涤、自动表面整修和墙面油漆等。

### ■ 产品和服务

- 与特定工业部门一起工作,以帮助把环境考虑融入产品、工艺过程和系统的设计与再设计;

- 通过该组织网站传播其各种工业项目、防污案例和技术报告;

- 开发可用于评估各种替代技术和其他防污选择的工具。

### ■ 网站

该网站提供有关DfE的各种工业项目的广泛信息;存取DfE的各种评估工具;与EPA的绿色化学和对环境友好的采购计划链接;与其他DfE、工业生态学和防污网站链接。

### ■ 联系方式

Design for the Environment Program

Office of Pollution Prevention and Toxics

U. S. EPA

Arel Rios Building

1200 Pennsylvania Ave., NW

Washington, D. C. 20460

United States

Phone: 202. 260. 1023

Fax: 202. 260. 4659

## 12. World Business Council for Sustainable development (WBCSD)

### ■ 组织概况

可持续发展的世界事务理事会(WBCSD)是一个 120 多家国际公司的联合体。它提供一种有关众多可持续发展问题的经营前景。它的基本目的是鼓励在经营方面各种环境管理的高标准。它还指导研究、促进合作和出版信息,以帮助各家公司和各种行业最好地去管理可持续发展的问题。

### ■ 产品和服务

WBCSD 通过以下方式向其国际会员服务:产生许多有关生物多样性到生态-效率/度量标准等问题的报告,以关于可持续经营方法方面的信息,来帮助经营或非经营团体推进已提高了的环境标准。它还帮助公司和工业界进行环境领域的巨大改进和对社会问题的更好了解。WBCSD 每个季度也出版自己的“贸易与环境公报”。

### ■ 网站

WBCSD 的网站发布它所有当前的工作动态,包括各种报告、股民对话摘要、案例、教育活动和-一个互动讨论网页;提供与其他网站的链接和一份可获得的出版物清单。

>ABS<	丙烯腈/丁二烯/苯乙烯
>ABS-FR<	阻燃剂 ABS
>EP<	环氧树脂
>PA<	尼龙(聚酰胺)
>PA6<	尼龙
>PA66<	尼龙
>PBT<	聚丁烯 对苯二酸酯
>PC<	聚碳酸酯
>PE<	聚乙烯
>PE-LLD<	线性低密度聚乙烯
>PE-LMD<	低-中密度聚乙烯
>PE-HD<	高密度聚乙烯
>PET<	聚乙烯 对苯二酸酯
>PS<	聚苯乙烯
>PS-HI<	耐冲击聚苯乙烯
>PVC<	聚氯乙烯
>SAN<	苯乙烯/丙烯腈
>SI<	硅

若要标记含有一种混合或多种混和的聚合体制成的零件,则此聚合体的代码便由“-”号分开。例如,对于一种 ABS 和聚碳酸酯的混和,便可用

### ■ 联系方式

World Business Council for Sustainable Development

160 route de Florissant. CH-1231

Conches-Geneva. Switzerland

Phone: +41. 22. 839. 3100

Fax: +41. 22. 839. 3131

以上所提供的资料,由于世事的变迁,可能随之发生变化,因此仅供参考。

## 附录 7 塑料标记的标准

关于各种塑料品种的标记方法,国际标准化组织给出了推荐的标准(Standards for Marking Plastic Parts)。此处特摘录如下,以供参考:

1) 只要尺寸、几何形状和功能允许,所有的塑料件,不论外部的或内部的,都要作标记。

2) 只要求新零件要作标记。眼前用现有工具准备制造的零件则随意。

3) 要使用 ISO(国际标准化组织;International Standards Organization)的术语和符号。标识塑料品种的大写英文字母的代码,被置于“>”(大于)符号和“<”(小于)符号之间,根据 ISO/DIS 11469,这便于辨识塑料标记而成为塑料的标识码。经常使用的工程塑料的标记示例如下所示:

acrylonitrile/butadiene/styrene
flame retardant/ignition resistant ABS
epoxy
nylon (polyamide)
6nylon 6
66nylon 66
polybutylene terephthalate
polycarbonate
polyethylene
linear low-density polyethylene
low-medium density polyethylene
high-density polyethylene
polyethylene terephthalate
polystyrene
high-impact polystyrene
polyvinylchloride
styrene/acrylonitrile
silicone

>ABS+PC<

4) 填充剂和增强剂则用两个字符代表,且紧随其后用一数字值代表填充剂/增强剂的重量百分比。几种

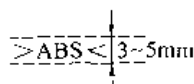
最常用的填充和增强材料的代码示例为:

GF	玻璃纤维	glass fiber
GB	玻璃珠	glass bead
MD	矿石粉	mineral powder
CF	碳纤维	carbon fiber

一种聚丁烯对苯二酸酯/聚碳酸酯的混和含有

>PC-GF20<  
 >(PBT+PC)-GF30<  
 >PC-(CF7+GF5-GB5)<  
 >(PE+PS)-(GF15+MD15)<

5) 下图中所示为推荐的字体高度是3~5mm, 而对于特大或特小的零件, 字体尺寸可以调整。只可以用大写字母。



6) 虽然不一定要在正常使用时所观察的部位上, 但是在每个产品上, 至少在一个外露零件的外边应出

### 参 考 文 献

- 1 铃木 胖编, リサイクル工学—循环型社会の构筑を目指して. 大阪: エネルギー 资源学会, 1996
- 2 熊文强, 郭孝菊, 洪卫编著. 绿色环保与清洁生产概论. 北京: 化学工业出版社, 2004
- 3 刘飞, 陈晓慧, 张华编著. 绿色制造. 北京: 中国经济出版社, 1999
- 4 刘光复, 刘志锋, 李钢编著. 绿色设计与绿色制造. 北京: 机械工业出版社, 1999
- 5 陶雪芬等编. 机械工业企业环境管理指南. 北京: 中国环境科学出版社, 1997
- 6 国家质量技术监督局发布. 中华人民共和国. GB/T24020-2000 idt ISO 14020:1998. 环境管理—环境标志和声明—通用原则
- 7 国家质量技术监督局发布. 中华人民共和国. GB/T24040-1999 idt ISO 14040:1997. 环境管理—生命周期评价—原则与框架
- 8 国家质量技术监督局发布. 中华人民共和国. GB/T24041-2000 idt ISO 14041:1998. 环境管理—生命周期评价—目的与范围的确定和清单分析
- 9 [日]山本良一编著. 环境材料. 北京: 化学工业出版社, 1997
- 10 蔡建国. 可持续发展战略与现代制造工业. 机电一体化. 1998(1)
- 11 [美]国防技术情报中心主编. 产品设计生产性指南. 张纯正, 张国梁等译. 北京: 航空工业出版

30%的玻璃纤维, 则可标记如下:

>(PBT+PC)-GF30<

注意, 在树脂与填充剂的代码之间是用“-”符号分隔的。还要注意, 在聚合物和填充剂的混合场合下要用适当的括号, 如以下示例所示:

一种聚合物, 一种填充剂  
 两种以上的聚合物, 一种填充剂  
 一种聚合物, 两种以上的填充剂  
 两种以上的聚合物, 两种以上的填充剂

现一个标记。例如, 一个合适的部位可以是紧挨着电气安全标记(UL), 或在其附近。这样显然无需首先拆解产品, 而其所有塑料零件已经有了回收标记。

7) 在零件的整个生命周期, 标记应保持清晰可读。在模塑零件上, 可以利用模具作出标记。其他可接受的方法有热冲压或烙印法。决不要用粘贴标签去标记塑料零件。

社, 1990

- 12 L. Alting, D. J. Jorgensen. The Life Cycle as a Basis for Sustainable Industrial Production. Annals of the CIRP, Vol. 42/1/1993
- 13 L. Alting, J. B. Legarth. Life Cycle Engineering Design. Annals of the CIRP, Vol. 44/2/1995
- 14 G. Boothroyd, et al. Design for Disassembly and Environment. Annals of the CIRP, Vol. 45/1/1996
- 15 S. L. Hart. Beyond Greening: Strategy for a Sustainable World. Harvard Business Review, 1997(Jan.-Feb.)
- 16 PRé Consulting. Eco-indicator 95
- 17 PRé Consulting. Eco-indicator 99
- 18 PRé Consulting. Impact Assessment
- 19 Business for Social Responsibility. White Paper: Green Product Design
- 20 D. F. Ciambone. Environmental Life Cycle Analysis. Lewis Publishers, Inc, 1997
- 21 M. A. Curran (ed.). Environmental Life Cycle Assessment. New York: McGraw-Hill, 1996
- 22 P. Hofstetter. Perspectives in Life Cycle Impact Assessment: A Structured Approach to combine Models of the Technosphere, Ecosphere, and Valuesphere. Kluwer Academic Publishers, 1998
- 23 United Nations Environmental Programme. Towards the Global Use of Life Cycle

- 
- Assessment. United Nations Publications, 2000
- 24 SETAC-Europe. Life Cycle Assessment. Brussels, Belgium, 1992
- 25 H. A. Udo de Haes(ed. ). Towards a Methodology for Life Cycle Impact Assessment. Brussels, Belgium;SETAC-Europe, 1996
- 26 L. Barnthouse. Life Cycle Impact Assessment: The State-of-the-Art 2<sup>nd</sup> edition. SETAC-Europe, 1998
- 27 E. S. Rubin, C. I. Davidson. Introduction to Engineering and the Environment. New York: McGraw-Hill, 2001

# 第 38 篇 智能设计

主 编 王安麟  
编写人 王安麟

# 第 1 章 智能模拟的科学

智能设计 (Intelligent Design ID) 主要通过智能模拟来实现。为此, 本章主要介绍思维科学的基础、思维的形式、智能模拟的方法以及智能模拟的神经基础和哲学基础。

## 1 信息社会与思维科学

随着生产自动化水平的不断提高和现代科学技术的迅猛发展, 人类社会已经进入了信息社会, 正在步入智能化的新时代。人们从来没有像今天这样重视信息在生产、生活、科研以及军事等方面的重要作用。由于人们面临着信息量大、传递迅速及复杂多变等特点, 因此, 对这些信息的获取、加工和处理变得更加困难和重要, 于是人们才真正感到, 要研究和利用人认识世界的规律和方法, 来提高人类自身的智能水平, 并使机器首先是计算机智能化, 就必须研究思维科学。

### 1.1 思维与思维科学

人们从不同角度研究思维的各个侧面已有悠久的历史。早在 20 世纪 80 年代初期, 我国著名科学家钱学森教授就倡导开展思维科学的研究。与此同时, 国外也开展了所谓认知科学 (Cognitive Science) 的研究, 它主要分为认识心理学和人工智能两个领域, 前者主要研究如何利用计算机仿真技术建立人的认知模型, 后者侧重如何运用人的认识经验使机器, 首先是计算机智能化。

国外的认知科学研究不涉及思维类型的基础理论研究, 只重视从个体角度研究思维, 尤其是尚未注重对形象思维机制的研究, 因此被看作是狭义的思维科学。

思维是人脑对客观事物间接的反映过程。所谓间接的反映, 意味着思维不是凭感觉器官对事物表象的直接认识, 而是通过间接的甚至迂回的途径来反映客观事物的特点或它们之间的联系与规律。间接认识需要借助于已有的知识和经验, 要间接地认识事物的特点、本质和规律, 绝不可能靠消极、被动地反映事物的表面现象, 必须靠自觉地、主动地在实践活动中占有材料, 靠回忆有关的知识 and 经验或通过联想、推想、想象等对有关材料进行分析、综合, “去粗取精、去伪存真、由此及彼、由表及里”地加工改造, 才能把握事物的本质, 找出事物间的规律性联系, 并有效地去改造客观事物。

人脑对客观事物的间接反映过程, 包括回想、联

想、表象、想象、思考、推想等。人们通过思维活动能够反映客观事物的特点、本质属性、内部联系及发展规律, 因此, 思维是认识过程的高级阶段。

意识是人的 一种认识活动, 它包括感觉、知觉和思维。思维是意识的一部分, 而且是最主要的成分, 假如没有思维, 人就不会有意识。思维科学是研究思维的规律和方法的科学, 而不涉及对具体思维内容的研究。思维科学的基础科学是研究人有意识思维规律的科学, 又称思维学。人的思维除了自己能够控制的意识以外, 还有很多人脑不能直接控制的意识, 即所谓的下意识。例如人走路开步走是人脑控制的, 走了二三步后就“自动化”了, 脑子并不去想该怎么走。要拐弯或遇到障碍时, 又控制一下, 所以, 人确实有很多意识是没有经过大脑的, 思维科学就是要研究人能够控制的那部分意识。

### 1.2 思维的类型

按照科学研究工作的需要, 从思维规律的角度出发, 思维可划分为抽象思维、形象思维和灵感思维三种类型。但是, 人的思维活动过程往往不是一种思维方式在起作用, 而是两种甚至三种先后交错在起作用。比如, 人的创造性思维的过程就决不是单纯的抽象思维, 总要包含点形象思维, 甚至要有灵感思维充当创造性思维火花的导火线和催化剂的作用。

#### 1.2.1 抽象 (逻辑) 思维学

抽象思维学又称逻辑思维学, 这里讲的逻辑是指人的思维规律。逻辑学分为形式逻辑和辩证逻辑两大类。

##### (1) 形式逻辑

形式逻辑是研究人们思维形式的结构及思维的基本规律的科学。思维形式的结构不是我们头脑中虚构出来的东西, 而是客观现实的一种反映, 它是客观事物的某种一般关系、特性的概括反映。如“S 是 P”这个结构, 它是“事物具有属性”这样一个事物的普遍性的反映; “M 是 P, S 是 M, 所以 S 是 P”是“全类是什么, 则全类事物中的一部分也就是什么”这样一种客观关系的反映。这种思维形式的结构是人类在长期实践活动中总结出来的产物。

思维的基本规律是运用各种思维形式时都必须遵守的规律。早在公元前 4 世纪, 已经有希腊哲学家亚里

上多德(B. C. Aristotte, 公元前384-322)创立了形式逻辑思维规律,即同一律、矛盾律和排中律。后来,到了17世纪末,德国哲学家莱布尼兹(G. W. Leibniz, 1646-1716)又增入了一条充足理由律,即组成了所谓的逻辑思维的四个初步规律。

从莱布尼兹开始,不少科学家和哲学家,特别是布尔(G. Boole, 1815-1864)和罗素(B. Russell, 1872-1970),把数学方法用于逻辑的研究,形成了数理逻辑这一学科,它可以看作是形式逻辑的一个特殊的分支。模态逻辑、多值逻辑、时序逻辑、模糊逻辑等,都属于数理逻辑这一范畴。形式逻辑又称传统逻辑,可以简称为逻辑。形式逻辑归根到底要解决的是思维的准确性问题。

### (2) 思维形式

思维的形式就是概念、判断和推理。概念是对客观事物的本质属性加以反映的思维形式。自然界及社会现象中的一切事物与现象,都具有许多性质。所谓性质就是事物所具有的那些相互区别、相互类似的一切质的、量的规定性,诸如数目、大小、速度、程度、动作、形态、特征、规律及关系等,都叫性质,它们是属于事物的,又称为事物的属性。属性可以分为本质的和非本质的两种。本质属性具有两个特点:其一,它是一个或一类事物内部所固有的规定性;其二,它具有把此事物和其他事物区别开的性质。本质属性一定是事物特有属性,而事物的特有属性却有一定是事物的本质属性。作为思维形式之一的概念,所反映的是事物的本质属性。

概念包含两个方面,一是概念的内涵,一是它的外延。一个概念的内涵就是这个概念对象的本质属性,而它的外延就是这个概念所反映的全体对象。概念的内涵和外延是概念两个有机联系方面,内涵是指外延对象的属性,外延是指具有内涵属性的对象。概念外延所构成的类就叫做集合。由此可见,研究集合就是从外延方面研究概念。逻辑学指出概念具有外延和内涵两个方面,为我们指出了一条明确概念、研究概念的途径;概念一样,都是思维形式的一种。判断是概念与概念的联合,而推理则是判断与判断的联合。在普通逻辑中,判断是对思维对象有所断定的思想,即断定对象具有某种属性,或不具有某种属性;断定的结果是肯定或否定某种对象及其属性。这是判断的最基本的逻辑特征。如果断定的情况被实践证明是符合客观实际的,那么这个判断就是真的,否则就是假的。因此,任何判断都是真的或者是假的,这种或真或假的性质叫做判断的值,这是判断的又一个基本特征。

推理是根据一个或一些判断获得一个新的判断的思维方式,任何推理都必须包含前提和结论两个组成

部分。已有的一个(或一些)判断称之为前提,新的判断称之为结论。在形式逻辑中,推理可以从不同角度分成多种形式,如直接推理、间接推理、演绎推理、归纳推理、类比推理、模态推理等。直接推理是指从一个前提推出结论的推理;间接推理是从两个或两个以上前提推出结论的推理;演绎推理是以一般的原理原则为前提,推到某个特殊的场合作出结论的推理方法;归纳推理是从若干个特殊的场合中的情况为前提,推求到一个一般的原理原则作为结论的推理方式;类比推理是由特殊性判断为前提,推出另一个特殊性的判断的推理方法;模态推理是最少有一个前提是模态判断的推理,所谓模态判断是对事物情况的性质加以判定的判断。

在科学论著中最常用的是演绎推理,它又分为三段论法和假言直言推理。三段论法是从两个判断(其中一个一定是“所有的S都是或不是P”的形式,S表示对象,P表示对象所具有的某种属性)得出第三个判断的推理方法。三段论包含着三个判断:第一个判断提供了一般的原理原则,称其为大前提;第二个判断指出了特殊场合的情况,叫小前提;联合这两个判断,说明一般原则和特殊情况间的联系,从而得出第三个判断,称之为结论。

假言推理和直言推理都属于演绎推理,假言推理的大前提是假言判断(是指肯定或否定对象在一定条件下具有某种属性的判断),小前提和结论是直言判断(无条件地肯定或否定某种事物的判断)。

### (3) 思维的基本规律

从上面论述的概念、判断和推理的思维形式可以看出,人们是按照一定的规律和逻辑结构去组织思想和进行思维的。逻辑的基本规律是客观事物的相对稳定性在思维中的反映,逻辑规律只是在思维活动中起作用,不在客观事物中起作用。事物的相对稳定性反映在思维中成为思维的确定性。思维的确定性表现为概念、判断的自身统一,这就是同一律;思维的确定性表现为概念和判断的前后一贯,不自相矛盾,这就是矛盾律;思维的确定性表现为两个相互矛盾的思想之间要作出抉择,排除中间的可能性,这就是排中律。

同一律是指在同一思维过程中,每一思想的自身都具有统一性,所谓思想的统一性是指概念或判断内容的同一性。科学研究的实践表明,任何一个严密完整的科学体系都是符合同一律要求的,如果违反了同一律的要求,科学研究就不能建立严密、完整的科学体系。

矛盾律是指在同一思维过程中,每个思想与其否定都不能同时为真,其中必有一假。矛盾律是把同一律思想进一步展开,指出既肯定又否定的思想是逻辑矛

盾,不能同真。矛盾律是用否定形式表示同一律用肯定形式表示的思想。矛盾律的作用是使思维首尾一致,避免自相矛盾。任何一个科学理论都具有不矛盾性,一个科学理论,如果包含有逻辑矛盾,人们就会对它发生怀疑。在科学史上,许多科学上的突破,往往是从发现原有科学体系的逻辑矛盾并在设法消除这种矛盾的基础上,创立了新的理论体系。

排中律是指在同一思维过程中,两个互相矛盾的思想必有一个是真的。排中律又比矛盾律深入一层,明确指出两个矛盾思想不能同假,必有一真。在论证中,矛盾律只能由真推假,不能由假推真,而排中律不是由真推假,而是由假推真。

同一律、矛盾律和排中律都是思维确定性的表现,它们之间的关系是密切的,只不过是不同侧面表述思维的确定性。它们构成了逻辑思维的基本规律,所有正确的思维形式都是以这些思维的基本规律作为基础。

#### (4) 辩证逻辑和数理逻辑

把高等数学关于变量等概念引进形式逻辑,促成了辩证逻辑的产生。辩证逻辑是关于思维运动的辩证规律的理论。

虽然形式逻辑和辩证逻辑都是研究思维形式,但是它们是从不同角度出发的。形式逻辑从抽象同一性角度研究思维形式,即把思维形式看作既成的相对稳定的范畴;辩证逻辑从具体同一性角度研究思维形式,即把思维形式看作对立统一、矛盾运动和转化的范畴。其次,形式逻辑的基本规律是同一律、矛盾律和排中律,它们虽有客观基础,但不是事物本质的规律;辩证逻辑的基本规律是对立统一、质量互变、否定之否定等规律,它们是客观事物本身的规律。

数理逻辑亦称符号逻辑,它源于形式逻辑,现已成为独立学科。数理逻辑是用数学方法研究推理、证明等问题的科学,主要内容为命题演算、谓词演算、递归论、证明论、集合论和模型论等。在形式化方面数理逻辑比形式逻辑更丰富、更发展。它用符号把概念、命题(判断)抽象为公式,把命题间的推理抽象为公式间的关系,并把推理转化为公式的推演。在数理逻辑中,用符号表示逻辑概念及其关系,常用符号“ $\rightarrow$ ”表示蕴含;“ $\neg$ ”表示否定;“ $\vee$ ”表示命题的析取(或);“ $\wedge$ ”表示合取(与);“ $\leftrightarrow$ ”表示等价。

数理逻辑关于形式语言的研究,为计算机语言提供了前提,而数理逻辑在计算机中的应用又推动了逻辑学的发展。辩证逻辑的建立和发展,对于提高人们的认识能力和推动形式逻辑、数理逻辑和电子计算机技术的发展具有非常重要的意义。

#### (5) 模糊逻辑和可拓逻辑

模糊逻辑是以模糊集合论为基础的,而传统的逻辑是以经典集合论为基础,通常称为二值逻辑,这种逻辑可以表述思维的确定性,但是它不能表述思维的模糊性。为了描述客观事物的模糊概念,美国加利福尼亚大学控制论专家扎德(L. A. Zadeh)教授在1965年发表了“模糊集合”的重要论文,从而创立了模糊数学。1975年扎德教授又出版了《模糊集合、语言变量及模糊逻辑》一书,标志着模糊逻辑的正式诞生。

在模糊逻辑中,将逻辑真值从普通的二值逻辑真值 $\{0, 1\}$ 扩展到了 $[0, 1]$ ,由于模糊逻辑真值在区间 $[0, 1]$ 中连续取值,通过该真值的大小表明真的程度。因此,模糊逻辑实质上是无限多值逻辑,也就是连续值逻辑,它为描述模糊概念及模拟人的模糊逻辑思维方式提供了强有力的工具。

将辩证逻辑和形式逻辑相结合产生了一种新的逻辑——可拓逻辑,它是以可拓集合为基础的。我国蔡文教授1983年发表了“可拓集合和不相容问题”的创见论文,目的在于研究解决现实世界中存在的不相容问题的规律和建立解决不相容问题的数学模型。论文中指出解决不相容问题,要考虑三个方面:一是必须涉及事物的变化及其特征;二是必须使用一些非数学方法;三是必须建立容许一定矛盾前提的逻辑。为此,建立了可拓集合的概念,以便讨论对象集内不属于经典子集而能转化到该子集内的元素,这是解决不相容问题的基础。在逻辑关系上,与可拓集合相对应,建立了关联函数的概念,把逻辑真值从 $(0, 1)$ 扩展到 $(-\infty, +\infty)$ ,用关联函数值的大小来衡量元素与集合的关系,使经典数学中“属于”和“不属于”集合的定性描述扩展为定量描述,以表征元素间的层次关系。

可拓逻辑能够描述事物的可变性,为解决客观世界中的矛盾问题提供了重要工具,它将在许多工程领域,如识别、决策、评价、控制、信息处理等方面有着广阔的应用前景。

### 1.2.2 形象(直觉)思维学

#### (1) 形象思维及其特点

形象思维,简单地说,就是凭借形象的思维。这种思维活动通过形象来思考和表述,它的主要思维手段是图形、行为等典型形象材料,它的认识特点是以个别表现一般,始终保留着事物的直观性,要求鲜明生动。思维过程主要表现为类比、联想、想象。

人的感觉器官接触到外界事物,通过大脑产生感觉,不同的感觉(视觉、听觉等)相互联系,经过综合以后形成知觉,知觉在脑中形成外界事物的感性形象,叫做映象,或称通过感性认识获得的表象,用表象进行的思维活动叫做形象思维,又称直觉思维。



表象是回想起过去感知过事物形象的过程。表象与感觉、知觉都是对事物外部形象的反映,但两者不同的是:感觉是对当前事物的直接反映,是由事物直接作用于感觉器官引起的,是认识事物的初级阶段;表象不是对当前事物的直接反映,而是对过去感知过的形象的再现。表象过程具有生动具体的形象,但表象过程不如知觉过程鲜明、完整和稳定。此外,表象具有间接的特征,它不是当前事物的直观形象,而是通过回想或联想在头脑中呈现的过去感知过的事物的形象。

概括地说,形象思维是在实践活动和感性经验基础上,以观念性形象即表象为形式,借助各种图式语音或符号语言为工具,以在经验中积累起来的形象知识为中介反映事物本质和联系的过程。

### (2) 形象思维的规律

**转换关联律:**在形象思维过程中,人们把事物的表象以及表象过程的信息转化成事物的状态信息,即通过表象反映事物的内在性质、内部变化和关系,必须事先在实践活动中建立起表象信息和状态信息的并联系统。比如,内科医生通过听诊器捕捉患者心肺活动的声象信息,然后把它转换成患者心肺状态的信息,这种信息转换的基础是医生头脑中建立了心肺声象和心肺状态信息的并联系统。

由于形象思维最基本的过程是形象信息与状态信息转换的过程,所以转换并联是形象思维的一条基本定律。

**模式补形律:**模式补形律是利用观念性的形象模式对事物或事物过程的表象进行整合补形,从而推出事物的补形或全形的规律。所谓观念性的形象模式,是指事物或事物过程的概括表象,是在长期实践过程中逐渐形成,它是对事物或事物过程的丰富形象特征进行分析、选择、概括、定型的结果,是形象思维中进行模式补形的内在根据。所谓整合补形是对事物不完整的、片面的表象进行加工、整理,同时补出缺少部分形象或补出事物完整形象的过程,它是一种形象思维的推理形式。

模式补形最主要的环节是建立事物的表象模式。在工程设计中,工程师把物体的形象抽象出来加以规范,采用简捷的线条表现出来,从而为施工人员提供了一个表象模式;在科学研究中,科学工作者对所研究的对象进行系统的研究,科学地确定每种对象的形象特征,于是就形成对象的表象模式。通过表象模式对事物不完整的形象进行整合补形是人类特有的一种形象思维能力,模式补形律是形象思维的一个普遍规律。

### (3) 形象思维的主要形式

形象思维的过程主要表现为类比、联想和想象。类比是通过两个不同对象进行比较的方法进行推

理,而重要的一环就是要找到合适的类比对象,这就要运用想象。类比方法在维纳控制论的形成和创立过程中起到关键的作用,正是采用类比沟通了机器、生命体和社会等性质不同的系统,找到了它们的相似性,为功能模拟方法的运用提供了逻辑基础。

联想是一种把工程技术领域里的某个现象与其他领域里的事物联系起来加以思考的方法。联想能够克服两个概念在意义上的差距把它们联系起来,联想的生理和心理机制是暂时的神经联系,也就是神经元模型之间的暂时联想。维纳就是利用类比和联想的方法,考究反馈在各种不同系统(从人的神经系统到技术领域)的表现,为控制论的形成奠定了基础。

想象是对头脑中已有的表象进行加工改造而创造新形象的思维过程。因此,它可以说是一种创造性的形象思维。想象不是直接感知过的事物的简单再现,而是对已有的表象进行加工改组形成新形象的过程。任何想象都必须以表象为基础,想象与表象既有区别又有联系,表象是现成的、旧有的,而想象是创造新形象的过程。想象的形成过程主要是对表象进行分析综合的加工改组过程,想象的分析综合是凭借形象来实现的。

想象对新知识的探索和科学发现具有重要作用,爱因斯坦曾说:“想象力比知识更重要,因为知识是有限的,而想象力概括着世界上的一切,推动着进步,并且是知识进化的源泉。严格地说,想象力是科学研究中的实在因素。”

著名的科学家钱学森指出:“人认识客观世界首先是用形象思维,而不是用抽象思维。就是说,人类思维的发展是从具体到抽象。”他建议把形象思维作为思维科学的突破口。因为它一旦搞清楚之后,就把前科学的那一部分,别人很难学到的那些科学以前的知识,即精神财富,都可以挖掘出来,这将把我们的智力开发大大地向前推进

### 1.2.3 灵感(顿悟)思维学

灵感思维是指人们在研究过程中对于曾经长期反复进行过探索而尚未解决的问题,因某种偶然因素的激发而豁然开朗,使其得到突发性顿悟的思维活动。灵感思维与直觉思维有某些相似之处,它们最主要的特点是产生突发性或偶然性。既突如其来,又稍纵即逝。在科学研究中,“灵机一动,计上心来”,也是这种灵感思维的表述。

灵感与机遇都同属一种偶然性,但二者性质又不相同,机遇发生在观察和实验中,属于客观现象,而灵感却产生于思考问题的过程中,属于主观现象。在科学史上,因偶然因素而产生灵感的事例是不胜枚举的。

钱学森指出：“如果逻辑思维是线性的，形象思维是二维的，那么灵感思维好像是三维的”，“研究人类的潜意识活动是搞清灵感思维机理的起步方向”。物质世界是一个三维的立体系统，物质世界的最高产物——人脑也是一个三维的立体系统。人脑不仅在意识这个呈现层次上反映立体的客观世界，而且在潜意识这个层次上反映立体的客观世界。潜意识(unconscious)是一个外来语，也译为无意识或下意识。所谓潜意识就是未呈现的意识，是人脑所具备的潜在的反映形式。潜意识的反映既不是人脑中固有的，也不是没有客观来源的，而是大脑这种特别复杂的物质机能，它是以一定的客体为对象的。现代实验心理学通过对脑阈下的各种不同的潜意识信息的电反应(诱发电位)的测定表明，它是客观存在着的。

灵感是人脑中显意识与潜意识交互作用而相互交融的结晶。然而，灵感思维的发生也有一个过程，在潜意识萌发酝酿灵感时，除潜意识推论外，还常有显意识功能的通力合作，当酝酿成熟时，突然与显意识沟通而涌现出来成为灵感思维。所谓潜意识推论是一种特殊的非逻辑性认识活动，它是多因素、多层次、多功能的系统整合过程。灵感思维实际上是一种潜意识思维方式，即是一种非逻辑思维，它同抽象思维、形象思维一样，都是人们理性认识所具备的一种高级认识方式。

灵感思维的基本特征是它的突发性、偶然性、独创性和模糊性，这些特征是它区别于其他思维形式的显著标志。

## 2 思维的基础和认知的发展

### 2.1 思维与智能

人类的智慧和才能是任何其他动物无法比拟的。人何以有这样高超的技能和本领呢？最根本的原因就是人类有比其他动物所不具有的发达的大脑。大脑是一切智慧行为的物质基础，没有高度发达的大脑，就不会有人类的智慧和才能，自然更谈不上发明与创造。所以，人类的本质特征就在于具有能够高度发展的智能。

一般说来，智能是指人类所特有的智慧和才能的综合。智慧是指辨析判断、发明创造的能力，才能是指知识和能力。知识是指人们在改造世界的实践中所获得的认识和经验的总和。能力可以理解为能胜任某项任务的主观条件。智力是智能的近义词，是指人认识、理解客观事物并运用知识、经验等解决问题的能力，包括观察、记忆、思维、想象、联想、判断、推理、决策等能力。所以，智能是人类所特有的智慧和能力的综合。概括说来，人类的智能就是人类认识世界和改造世界(包括自己在内)的才智(即才能和智慧)和本领。

人类智能的特点主要是思想，而思想的核心又是思维。可以说，没有思维就没有人类的智能，正是因为有了思维，人类的智能才能以远远超出动物而产生质的飞跃，出现了思想、意识，才使人类成为万物之灵。

### 2.2 思维的神经基础

人之所以能够感知和理解客观事物，作出反应，形成复杂的智能活动，是因为人具有产生这些心理活动的物质基础，那就是人类具有特殊的智能器官，主要是指高度发达的大脑，以及感觉器官、动作器官和语言器官。而在这些智能器官中最重要的是作为思维器官的大脑，它是人类产生智能并发展智能的物质基础。

19世纪末20世纪初自然科学开始研究人脑思维的生理机制。谢切诺夫(И. М. Сеченов, 1829—1905)的神经生理学和巴甫洛夫(И. П. Павлов, 1849—1936)的高级神经活动学说，揭示了人们的思维活动和大脑物质活动的生理过程的内在联系，证明大脑是人的思维器官，是人的智能活动的物质基础。

从20世纪中叶开始，一些心理学家、分子生物学家和人工智能专家，在神经生理学、高级神经活动学说的基础上，利用电子技术成果，从结构和功能方面进一步研究人脑的智能活动。研究结果表明，人的大脑由两半球组成，这两半球又由胼胝体联系起来。两半球的功能是有差异的，而两半球的功能由胼胝体加以协调。1981年美国著名神经生理学家诺贝尔奖金获得者斯佩里(R. W. Sperry, 1913)通过对裂脑人的精细实验证明，人脑的左半球主要同抽象思维、象征性关系和对象的细节的逻辑分析有关，它的主要功能是逻辑分析，体现了人的认识及有意识的行为，主要表现为划顺序的、分析的、语言的、局部的、线性的等特点；右半球与知觉、直觉和空间有关，具有形象思维的功能，具有音乐、绘画、综合、整体和几何空间的鉴别能力，主要表现为并行的、综合的、视觉空间的、非词语的、总体的、立体的等特点。

人类的高级行为首先应基于知觉，然后才能通过理性分析取得结果，这样的思维过程首先是由大脑左半球进行逻辑思维，然后通过右半球进行形象思维，再通过胼胝体联系并加以协调两半球的思维活动。在正常情况下，两半球之间存在着极为密切的联系，因而形象思维与抽象思维这两种思维方式不是截然分开的，而是互相交织、互相补充和互相转化的，从而达到对客观世界的更完美、更本质的认识。

灵感思维的神经基础是什么？近来，科学家们的最新研究成果表明，灵感的秘密在脑电波上。人在觉醒状态下进行思维活动时，大脑中有两种脑波，一种叫 $\alpha$ 波，一种称作 $\beta$ 波。 $\alpha$ 波是有规则地调和振动，表明精

神集中,大脑中有许多神经回路投入协调一致的工作。与此相反, $\beta$ 波是不规则不调和的振动,表示大脑的活动分散,精神不集中。科学家们认为,当灵感出现时,脑电波中 $\alpha$ 波就占优势。此时大脑中的潜意识大门打开,大脑思维可以抓住潜意识中所储存的主观信息,使其上升到意识中来,这就产生了智慧火花一闪间的悟性——灵感。

## 2.3 认知发展

### 2.3.1 皮亚杰认知发展理论

#### (1) 皮亚杰认知发展阶段理论

皮亚杰将儿童认知发展划分为四个阶段:感知运动阶段(出生~2岁)、前运算阶段(2~7岁)、具体运算阶段(7~11、12岁)、形式运算阶段(11、12~15、16岁)。

在感知运动阶段,儿童的智力只限于感知运动,儿童主要是通过感知运动图式与外界发生相互作用。在前运算阶段,儿童的思维已经表现出了符号性的特点,他们已能通过表象、言语以及其他的符号形式来表征内心世界和外在世界。但是其思维仍然是直觉性的,而非逻辑性的,且具有明显的自我中心特征。在前运算阶段早期,儿童主要使用象征来表征世界,随着年龄增长,越来越多地使用符号来表征外部世界。在具体运算阶段,儿童的思维具有了明显的符号性和逻辑性,基本上克服了思维中的自我中心性。但是其思维活动,在很大程度上仍然局限于具体的事物以及过去的经验,缺乏抽象性。所获得的最大收获,是具有了心理操作能力,儿童可以应用这种心理操作去认识、表征和反映内、外部世界,使其认知活动更具深刻性、灵活性和广泛性。在形式运算阶段,儿童总体的思维特点是能够设定和检验假设,能监控和内省自己的思维活动,思维具有抽象性。个体的心理操作之间,构成了有层次的组织系统,出现了所谓的“对操作的操作”(operations on operations)能力,使其不但能够注意其结果,而且还能够主动地监控、调整 and 反省自己的思维过程,其抽象思维能力获得很大的提高。皮亚杰认为,儿童在经过前述四个连续的发展阶段之后,其智力水平就基本趋于成熟。

#### (2) 皮亚杰认知发展机制

皮亚杰从生物适应的角度对儿童认知发展的内部变化过程进行了系统的分析和论述,形成了很有特色的儿童认知发展机制理论。其核心是认为儿童认知的发展是通过动作所获得的对客体的适应而实现的,适应的本质在于主体能取得自身与环境间的平衡,而达到平衡的途径是同化和顺应,在同化和顺应的过程中,

主体的认知操作能力获得系统化(组织化)的发展。

同化(assimilation)、顺应(accommodation)、平衡(equilibrium)、适应(adaptation)及组织(organization)是皮亚杰认知发展机制理论中的重要概念。

同化是指主体将其所遇到的外界信息直接纳入自己现有的认知结构中去的过程。在这个过程中,虽然主体对自身的认知结构并未进行任何调整和改善,但是这仍然是一个主动的过程。主体对外界信息所做的不仅仅是感觉登记,还需要对这些信息进行某些调整和转换,以使其与主体当前的认知结构相匹配,便于接纳。顺应是指主体通过调节自己的认知结构,以使其与外界信息相适应的过程。这个过程,对主体的认知结构的发展具有十分积极的意义。平衡为个体保持认知结构处于一种稳定状态的内在倾向性,并且,这种倾向性是潜藏于个体发展背后的一种动力因素。同时,平衡也是一个动态的过程,平衡包含着同化和顺应两个方面,平衡体现了主观存在与客观存在之间的最充分的相互作用。个体认知发展的过程就是不断地取得主、客体之间协调一致的过程。

认知发展的具体机制:当处于平衡状态的主体遇到某种新的信息时,由于新的信息与原有的认知结构之间存在着差距,便会出现不平衡状态。主体试图克服这种不平衡状态,所采用的方法有三种:①忽略。当前的外界信息与主体现有的认知结构差距过大,以致于主体根本不可能对此作出任何反应时,采用忽略而恢复到原平衡状态。主体的原有图式不发生变化。②同化。主体只需对外界信息略作调整或不作任何调整,就可将此纳入原有的认知结构中去。通过同化,主体原有的认知结构不会改变,或只获得某些量上的扩展,而没有质的变化,仍然回到原有的平衡状态。③顺应。主体通过调节自己的认知结构,以一种正确的方式对外界信息进行反应的过程,这时,主体的认知结构发生了质的变化,进入到一种新的、更稳定的平衡状态,主体的认知能力跃向一个新的水平。

皮亚杰认为,儿童的认知能力就是通过这种不断地从平衡——不平衡——平衡的运动而获得发展的,并且,主体的认知结构通过不断的同化和顺应,从原来的较为分散的状态整合到更高级的、更有组织的状态,这个认知结构就是他在《结构主义》中所试图阐述的存在于物理、生物等不同领域内的“结构”,它是由具有整体性的若干转换规律组成的一个有自身调整性质的图式体系,其自身调整过程也就是同化、顺应、平衡的过程。所以,同化、顺应、平衡、适应、组织化的过程及相互间的作用构成了儿童认知发展的内在机制,此即皮亚杰的认知发展机制理论。

### 2.3.2 斯腾伯格的认知三元素理论

斯腾伯格的智力三元素理论包括：背景性理论 (contextual subtheory)、经验性理论 (experiential subtheory)、组合性理论 (componential subtheory)。

#### (1) 背景性理论

背景性理论的核心观点是认为个体的智力状况是受其生活的环境影响的，在一定程度上，是由环境所决定的。在这个观点的基础上，斯腾伯格把个体的智力过程具体化为适应、选择和形成。“适应”是指主体通过调节自己的行为以使其更好地与周围的环境相适应。当适应遇到困难时，主体就会挑选另外一种环境，在这个新环境中，可能会实现更好的适应，这个过程就是“选择”。当主体既不能通过改变自己的行为来适应旧的环境，也不能选择新的环境时，就会去改变原有的环境，形成一个适合自己的新环境，此即“形成”。

#### (2) 经验性理论

其核心内容是强调主体所具有的知识和经验对当前的智力活动具有影响作用。主体在完成一种认知任务时，都需要有主体过去经验的参与，而那些达到了自动化的智力行为基本上就代表了过去经验积累的极限程度。

#### (3) 组合性理论

组合性理论的核心是斯腾伯格关于认知结构的三成分模式。这三个成分是：元成分、操作成分、知识获取成分。元成分的功能是对认知活动进行控制和调节；操作成分的功能是完成解决问题的具体步骤，如编码、提取信息，比较信息等；知识获取成分的功能是从长时记忆中提取与当前问题有关的信息或搜集新的信息。这三个成分相互激活，相互作用。

### 2.3.3 信息加工理论

信息加工理论更加强调了解人们在解决问题过程中的微观机制，所以，使用这种方法能使研究者获得更多的关于个体认识发展的比较具体的解释。信息加工方法本身不含发展的因素，但是它把大脑的思维过程看作是一个信息加工的过程，摒弃了“一元论”中所坚持的从物理结构来研究解释人脑的智能方法，避开了“一元论”研究方法遇到的许多难以解释的困难，而是从一个较高的抽象层次上来研究，取得了许多有意义的成果。更重要的，纽维尔 (A. Newell) 和西蒙 (H. Simon) 提出的符号处理系统，认为人脑是一个基于符号的信息处理过程，这是和计算机的处理机制一样的，从而为计算机的智能研究奠定了理论基础，并使计算机智能研究得以得到巨大发展。当然，由于其没有考虑信息的发展特征和人脑的许多特征不相符合，如

经验的形成和在系统中的作用与表现，在信息加工理论中难以有效地描述和体现出来，以致于西蒙在1995年发表文章《Artificial intelligence: an empirical science》，讨论了信息的发展和应用即经验问题，最终认为人工智能是一门经验科学。

### 2.3.4 思维的瞬间达尔文进化机制理论

威廉·卡尔文在其书《HOW BRAINS THINK》中，为我们描绘了智力演化的图景。他认为，在人的脑中，存在着许多时空模式（当然这些时空模式是基于功能观点抽象归纳出来的），这些时空模式可以是躯体的各种运动序列，可以是感知单元，对于输入信息的反应，是各种相关时空模式的复制、竞争过程。竞争取胜的时空模式得以在思维过程存在，并可作为新的信息输入，展开新一轮的竞争，这样实现思维的转移。综合其特征，即认为思维是一个瞬间的达尔文进化过程。

在瞬间的达尔文进化理论中，有几个比较重要的概念：分团、排序和达尔文进化。我们的工作记忆单元容量是一定的，魔数  $7 \pm 2$  是心理学家 George Miller 于1956年给出的工作记忆容量。容量的单位是以有意义的单元而存在，并不是字符，所以有了记忆结构“团”的概念，如记忆中的区号“021”是上海的，“010”是北京，记为一个“团”，而不是记为0、2、1、0、1、0三位数字。“团”是人脑的功能结构化所反映的一个侧面，人们热衷于以结构化的方式把事物串在一起，这个特性远远超出了其他动物所建立的序列性（反射序列性），如把音符组成旋律，把步子组成舞蹈等等。有了“团”这一有意义的结构体的概念，就不难理解在工作记忆单元一定的情况下，人的设计过程中能够同时想到几个待选方案的原因，这里每个待选方案都是一个团，也是一种图式。当有信息输入时，什么有意义的“团”进入工作记忆单元，取决于“团”与当前信息的相关程度，依据相关程度的高低给各“团”排序，最终哪一个“团”排在第一位，即作为当前思维存在，这个过程是一个各“团”的达尔文进化的竞争过程，当然这个过程是瞬间完成的。哪一个“团”取胜，其与当前信息的联系将加强，类似于自然界中生物的适应性得到加强。此即思维的瞬间达尔文进化理论。

### 2.3.5 广义进化认知模式

赵南元在其《认知科学于广义进化论》一书中，为我们论述了广义进化认知模式。其主要的论点如下：

#### (1) 脑的认知过程是一个典型的“自我表述系统”

人的每一次思考都会对以后的思维方式产生一定的影响。“自我表述”是指系统自身不但能够具有浅层知识，而且能够产生用浅层知识来表达和深化的深层

知识,即系统是一个能够不断自我繁衍的系统,知识能够不断增加,能力不断增强。引进自我表述概念是为了研究一种具有高度复杂性的系统,这种系统的复杂不在于系统中元素数量的多少、关系的多重性,而在于系统的结构甚至工作原理在系统的工作过程中能够改变,不但保持一定的秩序,而且还具有某种不稳定的因素,时刻引入新的东西,即处于秩序与混沌边缘的复杂。社会结构的进化、文化的进化、人脑的认知过程、生物的进化、生物个体发育过程等等,这些都是“自我表述系统”。

#### (2) 广义进化论的软硬结构模型

在其书中,用“地形图”模型表现了在固定条件下参数的寻优过程,同时也说明了生物进化过程中的急变与缓变问题,该问题是由化石研究中化石所提供的进化过程不是渐变过程,而是稳定与跳跃的交替过程而引出来的;用“博弈论”模型表现了在固定条件下方法、战略行为方式乃至价值观的择优过程。从地形图模型的观点来看,博弈论是一种处理变化地形图的方法,而对于我们最关心的建立价值观体系的要求,博弈论则提供了一种具有严密数学背景的求取评价函数的重要方法。但是对寻优和择优之后的结果如何影响其后的条件变化,缺乏一个有效的深层模型。对于一个连续的自我表述系统,我们不仅应注意到系统瞬间过程,而且还要注意到系统的成长过程,就不能不对系统进行更进一步的分析。

达尔文进化论的核心机制是变异和选择,广义进化论在此基础上加入软硬结构的相互作用机制。对于一个自我表述系统,可以分为两个部分:一部分叫做软结构,是可变的部分;另一部分是硬结构,是不变的部分。软结构与硬结构之间相互作用,硬结构对软结构提供支持作用,软结构对硬结构的作用是建构。硬结构执行系统的“日常”功能,保证系统的生存(存在),软结构在硬结构的支持下对硬结构进行建构,使整个系统不断进化(演化)可用滑模工艺来说明建造人工建筑物时的自举过程,计算机系统开发过程中已具有的软件和硬件作为硬结构,开发人员为软结构的开发系统模型,用汇编语言缩写一个“小C”的可执行编译程序,然后用此小编译程序开发一个功能更强的编译系统,人类社会中官僚与思想家的不同角色等等例子来说明软硬结构模型。

通过对生物进化的软硬结构分析,可以对进化重演律作出更准确的描述。生物发育的各个阶段并不是进化过程中各个时期动物成体的形态,而是各时期中被硬化的发育工序的积累。几乎可以断定,发育过程中后期形成的性状,一定是几乎后期发明的产物。在一种生物所拥有的大量遗传基因中,各基因的重要性是不

同的,有些基因发生变化对生物的生存毫无影响,有的基因发生变化就会引起致命的效果,前者称为软基因,后者称为硬基因。还有一些基因发生变化时对生物的成活率产生影响,但是并非致命。可折合为不到一个的硬基因。这样可以建立基因的硬信息量的概念。对于已经硬化的发育工序起指令作用的基因通常是硬基因。在同种的生物个体中,硬基因的差别很小,而软基因可以有较大的个体差异。硬基因由于受到严格的选择压力,可以表现出长期稳定不变,而软基因则在发生变异时不被除掉,呈现出与时间成正比的变化。生物在由低级向高级的进化过程中,硬信息量只增不减,随硬信息量的增加,生物对于突然变异的承受能力必须下降。这个推论与生物学中的实际观察是相符的。

把此软硬结构模型用于人脑的认知过程,同样认为是思维能力提高的过程比思维过程本身更重要,因为我们把学习和创造这样的动态过程作为研究的重点,从而对认知过程给出了机械论的解释,以便于在机器上实现含有创造性的智能。

#### 2.3.6 复杂自适应系统

在控制领域中,自适应也得到研究,形成了自适应控制系统,但是它们的目标仅仅是考虑如何在系统和环境及其变化规律不确定时,通过自动调整系统的结构或参数来减少不确定性,达到改善系统的品质,即如何使系统保持一种稳定状态。但是对于认知发展,其自适应性研究的是新的性质和结构如模式是如何突现出来的,而不是有意保持某种已有的确定性性质,这就是复杂自适应系统所研究的内容。复杂自适应系统,包括诸如人脑、免疫系统、生态系统、细胞、蚁群及人类社会中的政党、组织等等。

Holland 从对学习、进化和适应性的考虑中提出了系统的 building block 概念,认为在一个环境中,通过学习而改变 building block 之间的连接结构就是适应性的机制。对人类心智和适应性的研究使 Holland 进一步发展他的遗传算法,遗传算法虽然大体上抓住了进化的本质,但是过于简单,没有包含产生适应性的智能体, Holland 注意到了单个个体有限行为模型对复杂自适应系统模拟的不足,因此为每个智能体引入内部模型,每个 Agent 作为 building block,根据自己的模型进行预测,通过经验学习(这种学习是 Hebb 式的),从而产生了著名的分类系统,使系统不仅具有开采式的学习,而且具有探险式的学习,这是一种对心智的描述。

一般来说,共同进化或者适应的结果往往是导致生物体基因的改变,但是对于一个生命较长、智能较高的个体,如人、一些高等动物或者一些社会性的组织,

更加重要的改变不是他们基因的改变,而是他们在成长过程中所体现出来的学习行为与学习所得,这是“现场认知(学习)”所持有的观点,也是认知发展一直所研究的。“现场认知”强调的是,如果要考察个体的行为,研究如何使个体的行为日益合理,不断发展,应该从这个个体所属的群体来认识,包含着许多心理学成分在内,而不是仅仅将个体置于某个物理环境中,让它们相互作用(有限的作用)。

### 2.3.7 认知发展总论

认知发展的侧重点不仅在于对思维过程进行研究,而且对思维能力的发展进行研究,皮亚杰的认知发展阶段理论为我们描述了儿童思维能力的不断发展的宏观模式,当然这种研究是基于临床的小样本研究,但是通过对不同年龄阶段认知能力的比较,为我们提供了一个带有普遍性的理论。其认知发展内在机制:模式的同化、顺应、平衡过程为我们提出了一种思维的微观机制,可以看作是对思维的其中一个环节的展开和论述,通过这个过程不断演化,使认知能力螺旋式增长,其研究不是直接针对人脑的物理层次,而是处于一个较高的抽象层次,用抽象的结构或图式作为理论的基础。

斯腾伯格的认知三元素理论,则对认知的相关元素进行了归纳,分为背景性理论、经验性理论和组合性理论三个元素,分别对应着系统的环境及内在(先天)决定元素、系统知识积累、系统工作过程三个方面。

信息加工理论尽管对认知的发展没有进行很多的研究,这也正是目前计算机之所以没有适应性和学习能力,因而没有创造性的主要原因,但是它对于计算机能不能思维、能不能具有智能这个根本性的问题提出了实现的可能性,在人工智能四十多年的发展过程中,信息加工理论是其中占统治地位的符合主义的基础,目前的机器学习研究以提高知识系统的适应性和学习能力,也仍然需要以信息加工理论为基础,当然还要补充其不足,对认知发展方面进行更进一步的研究。

威廉·卡尔文的思维的瞬间达尔文进化理论则为我们提供了另外一种思维的微观过程,当然其研究和皮亚杰的认知微观机制研究一样,是建立在模式、结构或图式的基础上的,如果其模式能够得到合适的描述,则模式的进化过程就具有较强的操作性,模式的同化、顺应与平衡由较为简单的模式实用性作为评价准则的进化过程来代替,避免了认知结构与外来信息是否适应的难以机械判别问题,因此,对于应用思维的瞬间达尔文进化理论来建立人工智能系统,其研究的重点主要是能否找到合适的模式表达方法,这也正是本课题所研究的。

广义进化认知模式为我们生动地描述了在认知过程中认知结构是如何不断发展变化的,软硬结构模型中硬结构为软结构提供支持,软结构在硬结构的支持下对硬结构进行建构,软硬的转换,代表着系统的发展与进步、系统能力的提高,这正是“自我表述系统”应该具有的特性。

复杂自适应系统的研究,是对一类系统的描述,应用到认知发展上,则是不仅要注意到群体的实现特性,如遗传算法所体现出来的优化问题高鲁棒性、智能性和隐含并行性等,而且还要注意到在共同进化的过程中个体的能力增长机制如何,它们之间的关系在总论中已有所论述。

此外,人们还对认知发展的形式化描述进行了某种程度的研究,如李未的论文应用公理系统工具,对知识的增长、更新及假说的进化过程作了一定程度的刻画,通过引出新假设、事实反驳、假说重构等定义推导认知进程。从一种缺省逻辑(DL)称作必要性假设的理论出发作出合理假设,形成相应的NL理论来描述顿悟认知模拟问题,其理论基础是李未的开放逻辑中所提出的认知进程理论。同样,把模糊数学与数理逻辑相结合,对思维活动也进行了描述。

综上所述,这些研究的对象都是高度发达的人脑功能,并且把研究的重点放于非常令人感兴趣的认知发展上,当然也有从人脑的结构进行研究的,其文献更是无以数计,所关心的,是人脑智能的本质(这是三大难题之一)。

## 3 智能模拟

人工智能是20世纪中期产生并正在迅速发展的新兴边缘学科。它是探索和模拟人的智能和思维过程的规律,并进而设计出类似人的某些智能自动机的科学。人工智能的创始人温斯顿(P. H. Winston)认为,人工智能的中心任务是研究如何使计算机去做那些过去只有靠人的智力才能做的工作。本节主要从智能控制的角度,介绍智能模拟的科学基础、哲学基础、基本途径等问题。

### 3.1 智能模拟的科学基础

为了使机器具有某种人的智能,就必须研究生物机体的控制系统和人的思维活动规律与生理机制。从维纳创立的控制论,到工程控制论、经济控制论、生物控制论,再发展到智能控制论,这表明控制论这门科学随着科学技术发展及社会进步,它研究对象的领域不断的扩大,同时控制系统本身的控制决策的智能水平也在不断的提高。当今的科学技术比起维纳创立控制论的20世纪40年代,已经取得了长足的进步,这就为



进行智能模拟创造了良好的条件。但是,应该看到,人脑的思维活动极其复杂,模拟人脑思维活动是一项系统工程,它需要众多学科通力合作,多途径攻关,才能使得人工智能的水平逐步向人的智能水平接近。

神经生理学揭示了大脑是在长期实践活动中形成的高度组织起来的中枢神经系统,它是思维的物质基础。控制论和系统论运用系统的方法,从功能上揭示了机器和生物有机体不同系统所具有的共同规律,从而为人们从功能上模拟人的思维活动奠定了理论基础。计算机科学的发展,尤其是智能计算机的研制为智能模拟提供了理论和技术手段。思维科学的基础研究为智能模拟揭示了思维的规律和方法。模糊数学的创立和发展为模拟人的模糊逻辑思维方式提供了工具。人工神经网络理论研究为从结构和功能上模拟人的智能提供了重要手段。

随着科学技术的迅猛发展,新的学科会不断产生,由于智能模拟的需要,许多新兴学科又会高度的综合,其结果必将把智能模拟推向更高的水平。

### 3.2 智能模拟的哲学基础

众所周知,世界的统一性在于它的物质性,世界上除了运动的物质,除了千差万别的物质形态之外,再没有别的东西了。大脑的出现,是生物长期进化的产物。人脑是在社会实践的基础上产生和发展起来的高度复杂的物质体系,是思维的器官,而思维是人类智能的核心,是人脑的机能和属性。因此,从理论上讲,可以用物质的运动来模拟人脑的思维活动。

世界上永恒运动着发展着的物质是由运动形式决定的,物质的多样性和运动的多样性是紧密联系着的。物质的各种运动形式不仅可以互相转化,而且还存在着一种包含关系,即高级运动形式包含着低级运动形式,复杂的运动形式包含着简单的运动形式。这种包含关系反映了多种运动形式从低级到高级的发展变化规律。人的思维活动是自然界发展到社会运动形式的产物,是一种高级运动形式,同样也包含着一系列的低级运动形式。因此,人们可以借助计算机及其必要的可以实现的各种运动形式,来模拟人的思维活动。

由于思维与物质在本质上是统一的,在规则上是一致的,在运动形式上是包含的,因此,可以用其他形式的物质运动来模拟人的思维活动,这就是智能之所以能够模拟的哲学基础。应该指出,人工智能的根本目的是用物化的智能延伸和扩展人脑和机体的某些功能,智能模拟的根本方法是功能模拟法,模拟和被模拟的两个系统在结构和在实际过程中允许不一样,而且也没有必要要求一样。所以,模拟是仿真,而不是原型,模拟是近似,而不是等同,在这个意义上讲,模拟智能

就是仿人智能。

### 3.3 智能模拟的基本途径

自古以来,人们一直试图用各种机器来代替人的部分脑力劳动,以提高人类征服自然改造自然的能力。从20世纪50年代起,世界上许多控制论专家、计算机科学家、心理学家、仿生学家等分别从不同角度探讨智能模拟问题,概括起来主要有三种途径。

#### 3.3.1 基于逻辑推理的智能模拟——符号主义 (Symbolism)

符号主义起源于20世纪50年代中期,由纽维尔(A. Newell)和西蒙(H. Simon)提出模拟人类求解问题的心理过程,形成了物理符号系统。最早提出人工智能的研究者们都是从分析人类的思维过程出发,通过表示概念的符号及各种逻辑运算符、函数、过程等处理关系,获得具化表达。因此,概念被看作是智能模拟的核心,而把代表概念的符号作为基本元素,符号之间须满足一定逻辑运算关系。问题求解的过程,这种推理过程又可以通过某种形式化的语言加以描述。

符号主义的基本出发点是把人类思维逻辑加以形式化,并用一阶谓词逻辑加以描述问题求解的思维过程,这种基于逻辑的智能模拟实质上是模拟人的逻辑思维,或者说是实现模拟人的左脑抽象逻辑思维功能。

传统的人工智能学者普遍认为,思维,即应用有用信息的意识活动,是人类智能的体现,并试图以现代的计算机技术去模拟实现。然而,传统的二值逻辑难以实现人们在思维过程中的模糊概念形式化,而模糊逻辑的创立为人们把模糊逻辑思维形式化提供了新的有效途径。

#### 3.3.2 基于神经网络的智能模拟——联接主义 (Connectionism)

联接主义的先驱是心理学家麦卡洛克(W. S. McCulloch)和数学家匹茨(W. Pitts 1909)。他们在1943年合作的论文“神经活动中内在意识的逻辑运算”中,提出了形式化神经元模型,并认为由简单神经元构成网络,原则上可以进行大量复杂的计算活动。因此,早期的联接主义观点在本质上与符号主义无太大的区别。

20世纪60年代以后,由于许多科学家的不懈努力,其中包括Rosellblatt、Widrow、Kohonen、Hopfield、Crossberg以及Andersson等人的卓越贡献,使得人工神经网络技术有了重要的突破,为实现联接主义的智能模拟创造了条件。

联接主义者从生物,尤其是人的大脑神经系统的

构造和功能出发,把人的智能归结为脑的高层神经网络活动的结果,认为智能活动是大量简单的神经细胞,通过复杂的相互联接成网络后并行运行的结果。

联接主义认为神经细胞不仅是大脑神经系统的基本单元,而且是行为反应的基本单元,故称为神经元。任何思维和认知功能都不是少数神经元决定的,而是通过大量突触互相动态联系着的众多神经元协同作用来完成。

基于神经网络智能模拟方法,是以工程技术手段模拟人脑神经网络的结构与功能为特征,通过大量的非线性并行处理器来模拟众多的人脑神经细胞,用处理器错综灵活的连接关系来模拟人脑神经细胞之间的突触行为。这种连接机制的模拟方法,在一定程度上有可能起到对人脑形象思维的模拟,即承担了人脑右半球形象思维功能的模拟。

### 3.3.3 基于“感知—行动”的智能模拟——行为主义 (Behaviourism)

早在1919年美国心理学家华特生(J. B. Watson, 1878—1958)就提出心理学应该是一门行为科学的观点,从而丰富和推动了心理学的研究和发展。1948年,维纳在著名的《控制论》一书中指出:控制论是在自控理论、统计信息理论和生物学的基础上发展起来的。它着重研究机器自适应、自组织、自修复和学习机理,这些功能由系统的输入输出反馈行为所决定。

在控制论发展的初始阶段,计算模型是模拟的,该领域的许多工作实际上是随着对动物和智能的了解,并希望探明动物如何通过学习来改变他们的行为,以及如何导致对整个环境的适应。早在1952年, Ashby曾指出:为了理解机体所产生的行为,一个机体和它周围的环境必须一起构成模型。著名心理学家皮亚杰曾指出:“逻辑的根源必须从动作(包括言语行为)的一般协调中去探求”,“逻辑数理运算来源于行动本身,因为它从行为的协调中抽象出来的结果,而不是从对象本身演绎出来的”。

美国麻省理工学院青年教授 Brooks 在1990、

1991年相继发表论文,提出无需表达和推理的智能行为的观点,从而成为人工智能研究中行为主义的杰出代表人物。

#### (1) 智能、思维与行为的关系

前面已经指出,模拟人的智能就是模拟人的思维形式,基于这样的观点实现智能模拟目前面临的主要困难表现在三个方面:第一,由于人脑的真实思维模型无法获得,因此在智能模拟系统中的抽象、表达、推理及学习等方法的正确性受到限制;第二,人脑的思维具有并行的特点,目前采用冯·诺伊曼式计算机无法实现模拟人脑并行思维的过程;第三,联接主义虽然具有并行性特点,但目前网络的优化拓扑结构及快速收敛性学习算法难以实现。

实际上,人的正确思维活动离不开实践活动,从广义上讲,人与环境的交互作用,体现了人的思维与感知—运动的行为之间的密切关系。根据存在决定意识及意识反作用(能动作用)的哲学观点,以及应用有用信息的意识活动,即是思维并体现智能的观点,人工智能还应该研究思维与行为的交互关系。

#### (2) Brooks 行为主义的观点

Brooks 对人工智能研究认为,首先要弄清楚生命系统在复杂的自然环境中所具有的生存和反应能力的本质,然后才有可能进一步探讨人类高水平的智能问题。这种观点的本质就是适应自然环境的感知—运动行为模式。

Brooks 在基于动作分解原理的动作理论指导下,完成了一个六足行走机器人实验系统,共包含150多个传感器和23个执行器。Brooks 动作理论的核心思想是动作分解,而不是传统的功能分解,这样就能用简单的有限状态机方法将感知器和执行器有机地集成,以形成行为产生器,即感知—运动模块。Brooks 的机器人实验系统在白然环境中所表现的防碰撞、漫游动作、行为的灵活性给人以深刻的印象。尽管这种人造生物的智能水平还处于仿昆虫的低级阶段,但是 Brooks 的基于行为的研究方法却为智能模拟提供了一个新的途径。



## 第2章 智能设计方法和技术综述

### 1 智能设计的发展概述

#### 1.1 CAD的发展

以依据算法的结构性能分析和计算机辅助绘图为主要特征的传统CAD技术在产品设计中的成功应用,引起设计领域内的一场深刻变革。包括设计活动在内的求解大致可分为两类工作:第一类是基于数学模型和数值处理的计算型工作;第二类是基于符号性知识模型和符号处理的推理型工作。传统CAD技术在数值计算和图形绘制上扩展了人的能力,可以比较圆满地完成第一类工作,但对第二类工作往往难以胜任。由于产品设计是人的创造力与环境条件交互作用的物化过程,是一种智能行为,通常需要设计人员分析推理、运筹决策和综合评价,才能取得合理的结果。为了对设计的全过程提供有效的计算机支持,传统CAD系统有必要扩展为智能CAD系统。

通常把提供了诸如推理、知识库管理查询机制等信息处理能力的系统定义为知识处理系统,例如专家系统就是一种知识处理系统。具有传统计算能力的CAD系统经这种知识处理技术加强后称为智能CAD(ICAD)系统<sup>[1]</sup>。ICAD系统把专家系统等人工智能技术与优化设计、有限元分析、计算机绘图等各种数值计算技术结合起来,取其所长,其目的是尽可能地使计算机参与方案决策、结构设计、性能分析和图形处理等设计全过程。因此,ICAD系统除了具有工程数据库、图形库等CAD功能模块外,还应具有知识库、推理机等智能模块。

虽然ICAD可以提供对整个设计过程的计算机支持,但完成第一类和第二类工作的功能模块是彼此分隔、松散耦合的,它们之间的连接仍然要由人类专家完成。近年来,随着高技术的发展和社会需求的多样化,小批量多品种生产方式的比重不断加大,CIMS应运而生并迅速发展。在CIMS这样的集成环境下,产品设计技术日趋复杂,已不可能也不允许将设计活动划分为计算型、推理型这样彼此分隔的独立结构。面向CIMS的设计活动既包括计算型、推理型工作,也包括其他性质类型的工作,如利用样本性知识进行自学习,而且在设计的每一阶段,各种不同性质类型的工作彼此交融,难以分离。这样,在CIMS技术的推动下,ICAD系统应在原有基础上强化集成功能,由此被提

升到一个新的阶段,即集成化智能CAD(I2CAD)阶段,它是面向CIMS的ICAD系统,可对设计全过程提供一体化的计算机支持。

#### 1.2 智能设计的两个阶段

智能设计的产生可以追溯到专家系统技术最初应用的时期,其初始形态都采用了单一知识领域的符号推理技术——设计型专家系统,这对于设计自动化技术从信息处理自动化走向知识处理自动化有着重要意义,但设计型专家系统仅仅是为解决设计中某些困难问题的局部需要而产生的,只是智能设计的初级阶段。

近10年来,CIMS的迅速发展向智能设计提出了新的挑战。在CIMS这样的环境下,产品设计作为企业生产的关键性环节,其重要性更加突出。为了从根本上强化企业对市场需求的快速反应能力和竞争能力,人们对设计自动化提出了更高的要求,在计算机提供知识处理自动化(这可由设计型专家系统完成)的基础上,实现决策自动化,即帮助人类设计专家在设计活动中进行决策。需要指出的是,这里所说的决策自动化决不是排斥人类专家的自动化。恰恰相反,在大规模的集成环境下,人在系统中扮演的角色将更加重要。人类专家将永远是系统中最有创造性的知识源和关键性的决策者。因此,CIMS这样的复杂巨系统必定是人机结合的集成化智能系统。与此相适应,面向CIMS的智能设计走向了智能设计的高级阶段——人机智能化设计系统。虽然它也需要采用专家系统技术,但只是将其作为自身的技术基础之一,与设计型专家系统之间存在着根本的区别。

设计型专家系统解决的核心问题是模式设计,方案设计可作为其典型代表。与设计型专家系统不同,人机智能化设计系统要解决的核心问题是创新设计,这是因为在CIMS这样的大规模知识集成环境中,设计活动涉及多领域和多学科的知识,其影响因素错综复杂。CIMS环境对设计活动的柔性提出了更高要求,很难抽象出有限的稳态模式。换言之,即使存在设计模式的话,设计模式也是千变万化,几乎难以穷尽。这样的设计活动必定更多地带有创新色彩,因此创新设计是人机智能化设计系统的核心所在。

设计型专家系统与人机智能化设计系统在内核上存在差异,由此可派生出两者在其他方面的不同点。例如,设计型专家系统一般只解决某一领域的特定问题,

比较孤立和封闭,难以与其他知识系统集成;而人机智能化设计系统面向整个设计过程,是一种开放的体系结构。

智能设计的发展与CAD的发展联系在一起,在CAD发展的不同阶段,设计活动中智能部分的承担者是不同的。传统CAD系统只能处理计算型工作,设计智能活动是由人类专家完成的。在ICAD阶段,智能活动由设计型专家系统完成,但由于采用单一领域符号推理技术的专家系统求解问题能力的局限,设计对象(产品)的规模和复杂性都受到限制,这样ICAD系统完成的产品设计主要还是常规设计,不过借助于计算机支持,设计的效率大大提高。而在面向CIMS的ICAD,即I2CAD阶段,由于集成化和开放性的要求,智能活动由人机共同承担,这就是人机智能化设计系统,它不仅可以胜任常规设计,而且还可支持创新设计。因此,人机智能化设计系统是针对大规模复杂产品设计的软件系统,它是面向集成的决策自动化,是高级的设计自动化。

## 2 智能设计的概念和特征

智能设计系统,应该不仅仅是对人脑某些思维特征(如抽象思维、形象思维)的模拟,而且应该具有自学习、自适应的能力,即具有自我进化的机制(进化智能),来保证系统的生命力及解决问题的有效性。面对21世纪产品竞争日益加剧的挑战,世界各国普遍重视提高产品的设计水平,以增强产品竞争力,同时,我国市场经济的发展对产品设计开发与开发提出了强烈的创新要求。在目前激烈的市场竞争中,对产品设计除了要求新颖独特和性价比高以外,还要求设计快捷方便和高效率。

分析现有的产品设计理论和方法,提高设计效率有以下两种策略,一种是对产品设计进程的动态重组,通过设计的组织形式的改变来尽量缩短设计周期,如目前的CIMS、并行工程和并行设计等,这些和管理科学结合的方法使工作更有效率,需注意的是,管理的操作主要是针对活动主体,如设计师等;另外一种策略是借助于计算机,通过计算机的高速、海量运算能力来缩短设计周期,取代人的部分体力和脑力劳动,CAD、智能CAD技术的引入就是基于这种策略,这些是计算机技术和工人智能的结合,操作对象是设计工具的智能化。这两种策略是互相交互的。

目前CAD技术和智能CAD技术发展还远未达到人们对它所抱的期望。首先是现有的CAD技术主要应用于设计过程的后期,如详细设计阶段、图形处理等,还没有进入设计的早期——概念设计阶段,这是由于在设计早期需要更多地用到创造性思维(包含了经

验思维、抽象逻辑思维和形象思维)。其次是目前的人工智能在抽象度高的方面以信息加工理论为理论基础,所构造的专家系统缺少适应性和灵活性,知识获取和进化成为知识系统的瓶颈;抽象度低的结构主义——人工神经网络是经验映射和形象映射的隐含表示,难以解释高层的思维活动,所以达到的智能化水平和人们的厚望相差甚远。

传统CAD系统由于缺乏设计工程师所具有的推理和决策能力,已经不能满足设计过程自动化的要求。而智能CAD(ICAD)系统既具有传统CAD系统的数值计算和图形处理能力,又有知识处理能力,能够对设计的全过程提供智能化的计算机支持,这就是对智能CAD理论和应用的研究。

### 2.1 智能设计的特点

1) 以设计方法学为指导。智能设计的发展,从根本上取决于对设计本质的理解。设计方法学对设计本质、过程设计思维特征及其方法学的深入研究是智能设计模拟人工设计的基本依据。

2) 以人工智能技术为实现手段。借助专家系统技术在知识处理上的强大功能,结合人工神经网络和机器学习技术,较好地支持设计过程自动化。

3) 以传统CAD技术为数值计算和图形处理工具,提供对设计对象的优化设计、有限元分析和图形显示输出上的支持。

4) 面向集成智能化。不但支持设计的全过程,而且考虑到与CAM的集成,提供统一的数据模型和数据交换接口。

5) 提供强大的人机交互功能。使设计师对智能设计过程的干预,即与人工智能融合成为可能。

### 2.2 智能设计技术的研究重点

1) 智能方案设计。方案设计是方案的产生和决策阶段,是最能体现设计智能化的阶段,是设计全过程智能化必须突破的难点。

2) 知识获取和处理技术。基于分布和并行思想的结构体系和机器学习模式的研究,基于基因遗传和神经网络推理的研究,其重点均在非归纳及非单调推理技术的深化等方面。

3) 面向CAD的设计理论。包括概念设计和虚拟现实、并行工程、健壮设计、集成化产品性能分类学及目录学、反向工程设计法及产品生命周期设计法等。

4) 面向制造的设计。以计算机为工具,建立用虚拟方法形成的趋近于实际的设计和制造环境。具体研究CAD集成、虚拟现实、并行及分布式CAD/CAM系统及其应用、多学科协同、快速原型生成和生产的设计

等人机智能化设计系统 (I2CAD)。智能设计是智能工程与设计理论相结合的产物,它的发展必然与智能工程和设计理论的发展密切相关,相辅相成。设计理论和智能工程技术是智能设计的知识基础。智能设计的发展和实践,既证明和巩固了设计理论研究的成果,又不断提出新的问题,产生新的研究方向,反过来还会推动设计理论和智能工程研究的进一步发展。智能设计作为面向应用的技术其研究成果最后还要体现在系统建模和支撑软件开发及应用上。

## 2.3 智能化方法的分类和智能设计的层次

### 2.3.1 智能化方法的分类

智能设计归根结底是要在设计过程中模拟人的智能的决策方式,模拟人的智能实质上是模拟人的思维方式。智能设计的基础是人工智能技术。尽管人工智能已经创造了一些实用系统,但人们不得不承认这些远未达到人类的智能水平。其方法可分为两大类:

#### (1) 第一类:自上而下的方式(符号处理的方法)

其是基于 Newell 和 Simon 的物理符号系统的假说。尽管不是所有人都赞同这一假说,但几乎大多数被称为“经典的人工智能”(即哲学家 John Haugeland 所谓的“出色的老式人工智能”或 GOFAI) 均在其指导之下。这类方法中,突出的方法是将逻辑操作应用于说明性知识库。这种风格的人工智能运用说明语句来表达问题域的“知识”,这些语句基于或实质上等同于一阶逻辑中的语句。采用逻辑推理可推导这种知识的结果,这种方法有许多变形,包括那些强调对逻辑语言中定义域的形式公理化的角色的变形。当遇到“真正的问题”,这一方法需要掌握问题域的足够知识,通常就称作基于知识的方法。在大多数符号处理方法中,对需求行为的分析和为完成这一行为所做的机器合成要经过几个阶段。最高阶段是知识阶段,机器所需知识在这里说明。接下来是符号阶段,知识在这里以符号组织表示(例如列表可用列表处理语言 LISP 来描述),同时在这里说明这些组织的操作。接着,在更低级的阶段里实施符号处理。多数符号处理采用自上而下的设计方法,从知识阶段向下到符号和实施阶段。

#### (2) 第二类:自下而上的方式(“子符号”方法)

其通常采用自下而上的方式,从最低阶段向上进行。在最低层阶段,符号的概念就不如信号这一概念确切了。在子符号方法中突出的方法是“Animat approach”。偏爱这种方式 (Wilson 1991, Brooks 1990) 的人们指出,人的智能经过了在地球上十亿年或更长时间的进化过程。他们认为,为了制造出真正的智能机器,必须沿着这些进化的步骤走。因此,必须集中

研究复制信号处理的能力和简单动物,如昆虫的支配系统,沿着进化的阶梯向上进行。这一方案不仅能在短期内创造实用的人造物,又能为更高级智能的建立打好坚实的基础。

第二类方法也强调符号基础。Brooks 1990 年将物理符号系统和他的物理基础假说相对照。在物理基础假说中,一个 agent 不采用集中式的模式,而运用其不同的行为模块与环境相互作用来进行复杂的行为(然而,他也承认,要达到人类智能水平的人工智能也许需要将两种途径相结合)。

机器与环境的相互作用产生了所谓的“自然行为(emergent behavior)”。一些研究人员认为,一个 agent 的功能可视为该系统与动态环境密切相互作用的自然属性。agent 本身对其行为的说明并不能解释它运行时所表现的功能;相反,其功能很大程度上取决于环境的特性。不仅要动态地考虑环境,而且环境的具体特征也要运用于整个系统之中。

由于符号派制造的著名样品机器包括所谓的“神经网络(Neural network)”。受到生物学方法的启发,这些系统主要因其学习的能力而十分有趣。根据模拟生物进化方面的进程,一些有趣的机器应运而生,包括:Sexual crossover, Mutation 和 Fitness-proportional reproduction。

介于自上而下和自下而上之间的方法是一种动机“环境自动机(situated automata)” (Kaelbling & Rosenschein 1990, Rosenschein & Kaelbling 1995) 的方法。Kaelbling 和 Rosenschein 建议编写一种程序设计语言来说明 agent 在高水平上所要求的行为,并编写一编译程序,以从这种语言编写的程序中产生引发行为的线路。

### 2.3.2 智能设计的层次

综合国内外关于智能设计的研究现状和发展趋势,智能设计按设计能力可以分为三个层次:常规设计、联想设计和进化设计。

#### (1) 常规设计

即设计属性、设计进程、设计策略已经规划好,智能系统在推理机的作用下,调用符号模型(如规则、语义网络、框架等)进行设计。目前,国内外投入应用的智能设计系统大多属于此类,如日本 NEC 公司用于 VLSI 产品布置设计的 Wirex 系统,华中理工大学开发的标准 V 带传动设计专家系统(JDDDES)、压力容器智能 CAD 系统等。这类智能系统常常只能解决定义良好、结构良好的常规问题,故称常规设计。

#### (2) 联想设计

目前研究可分为两类:一类是利用工程中已有的

设计事例,进行比较、获取现有设计的指导信息,这需要收集大量良好的、可对比的设计事例,对大多数问题是困难的;另一类是利用人工神经网络的数值处理能力,从试验数据、计算数据中获得关于设计的隐含知识,以指导设计。这类设计借助于其他事例和设计数据,实现了对常规设计的一定突破,称为联想设计

### (3) 进化设计

遗传算法(GA,即 Genetic algorithms)是一种借鉴生物界自然选择和自然进化机制的、高度并行的、随机的、自适应的搜索算法。20世纪80年代早期,遗传算法已在人工搜索、函数优化等方面得到广泛应用,并推广到计算机科学、机械工程等多个领域。进入20世纪90年代,遗传算法的研究在其基于种群进化的原理上,拓展出进化编程(EP,即 Evolutionary programming)、进化策略(ES,即 Evolutionary strategies)等方向,它们并称为进化计算(EC,即 Evolutionary computation)。

进化计算使得智能设计扩展到进化设计,其特点是:

1) 设计方案或设计策略编码为基因串,形成设计样本的基因种群。

2) 设计方案评价函数决定种群中样本的优劣和进化方向。

3) 进化过程就是样本的繁殖、交叉和变异等过程。

进化设计对环境知识依赖很少,而且优良样本的交叉、变异往往是设计创新的源泉,所以在1996年举办的“设计中的人工智能”(Artificial intelligence in design'96)国际会议上,M.A.Rosenman提出了设计中的进化模型,进而将进化计算作为实现非常规设计的有利工具。

综上所述,智能设计的研究随着人工神经网络、进化计算等技术的引入,处于由常规设计、联想设计向创新设计突破的关键阶段,有很多工作值得深入的研究和探讨。

## 2.4 智能设计的基本方法

### 2.4.1 智能设计的分类

#### (1) 原理方案智能设计

方案设计的结果将影响设计的全过程,对于降低成本、提高质量和缩短设计周期等有至关重要的作用。原理方案设计是寻求原理解的过程,是实现产品创新的关键。原理方案设计的过程是总功能分析——功能分解——功能元(分功能)求解——局部解法组合——评价决策——最佳原理方案。按照这种设计方法,原理方案设计的核心归结为面向分功能的原理求解。面向通用分功能的

设计目录能全面地描述分功能的要求和原理解,且隐含了从物理效应向原理解的映射,是智能原理方案设计系统的知识库初始文档。基于设计目录的方案设计智能系统,能够较好的实现概念设计的智能化。

#### (2) 协同求解

ICAD应具有多种知识表示模式、多种推理决策机制和多个专家系统协同求解的功能,同时需把同理论相关的基于知识程序和方法的模型组成一个协同求解系统,在元级系统推理及调度程序的控制下协同工作,共同解决复杂的设计问题。

某一环节单一专家系统求解问题的能力,与其他环节的协调性和适应性常受到很大限制。为了拓宽专家系统解决问题的领域,或使一些互相关联的领域能用同一个系统来求解,就产生了所谓协同式专家系统的概念。在这种系统中,有多个专家系统协同合作,这就是协同式多专家系统。多专家系统协同求解的关键,是要工程设计领域内的专家之间相互联系与合作,并以此来进行问题求解。协同求解过程中信息传递的一致性原则与评价策略,是判断目前所从事的工作是否向着有利于总目标的方向进行。多专家系统协同求解,除在此过程中实现并行特征外,尚需开发具有实用意义的多专家系统协同问题求解的软件环境。

#### (3) 知识获取、表达和专家系统技术

知识获取、表达和利用技术专家系统技术是ICAD的基础,其面向CAD应用的主要发展方向,可概括为:

1) 机器学习模式的研究,旨在解决知识获取、求精和结构化等问题。

2) 推理技术的深化,要有正、反向和双向推理流程控制模式的单调推理,又要把重点集中在非归纳、非单调和基于神经网络的推理等方面。

3) 综合的知识表达模式,即如何构造深层知识和浅层知识统一的多知识表结构。

4) 基于分布和并行思想求解结构体系的研究。

#### (4) 黑板结构模型

黑板结构模型侧重于对问题整体的描述以及知识或经验的继承。这种问题求解模型是把设计求解过程看作是先产生一些部分解,再由部分解组合出满意解的过程。其核心是由知识源、全局数据库和控制结构三部分组成。全局数据库是问题求解状态信息的存放处,即黑板,将解决问题所需的知识划分成若干知识源,它们之间相互独立,需通过黑板进行通信、合作并求出问题的解。通过知识源改变黑板的内容,从而导出问题的解。在问题求解过程中所产生的部分解全部记录在黑板上,各知识源之间的通信和交互只通过黑板进行,黑板是公共可访问的。控制结构则按人的要求控制知识

源与黑板之间的信息更换过程,选择执行相应的动作,完成设计问题的求解。黑板结构模型是一种通用的适于大空间解和复杂问题的求解模型。

#### (5) 基于实例的推理 (CBR)

CBR 是一种新的推理和自学习方法,其核心精神是用过去成功的实例和经验来解决新问题。研究表明,设计人员通常依据以前的设计经验来完成当前的设计任务,并不是每次都从头开始。CBR 的一般步骤为提出问题,找出相似实例,修改实例使之完全满足要求,将最终满意的方案作为新实例存入实例库中。CBR 中最重要的支持是实例库,关键是实例的高效提取。

CBR 的特点是对求解结果进行直接复用,而不用再次从头推导,从而提高了问题求解的效率。另外,过去求解成功或失败的经历可用于动态地指导当前的求解过程,并使之有效地取得成功,或使推理系统避免重犯已知的错误。

### 2.4.2 智能设计系统与技术

#### (1) 智能设计系统

在 CIMS 环境下,为了提高制造业对市场变化和小批量、多品种要求的迅速响应能力,设计正在向集成化、智能化、自动化方向发展。要实现这一目标,就必须大大加强设计专家与计算机工具这一人机结合的设计系统中机器的智能,使计算机能在更大范围内、更高水平上帮助或代替人类专家处理数据、信息与知识,做出各种设计决策,大幅度提高设计自动化的水平。智能设计就是要研究如何提高人机系统中计算机的智能水平,使计算机更多更好地承担设计中各种复杂任务,成为设计工程师得力的助手和同事。

在设计技术发展的不同阶段,设计活动中智能部分的承担者是不同的。以人工设计和传统 CAD 为代表的传统设计技术阶段,设计智能活动是由人类专家完成的。在以 ICAD 为代表的现代设计技术阶段,智能活动由设计型专家系统完成,但由于采用单一领域符号推理技术的专家系统求解问题能力的局限,设计对象(产品)的规模和复杂性都受到限制,不过借助于计算机支持,设计的效率大大提高,而在以 I2CAD 为代表的先进设计技术阶段,由于集成化和开放性的要求,智能活动由人机共同承担,这就是人机智能化设计系

统。虽然人机智能化设计系统也需要采用专家系统技术,但它只是将其作为自己的技术基础之一,两者仍有较根本的区别:

1) 设计型专家系统只处理单一领域知识的符号推理问题;而人机智能化设计系统则要处理多领域知识,多种描述形式的知识,是集成化的大规模知识处理环境。

2) 设计型专家系统一般只解决某一领域的特点问题,比较孤立和封闭,难以与其他知识系统集成;而人机智能化设计系统则面向整个设计过程,是一种开放的体系结构。

3) 设计型专家系统一般局限于单一知识领域范畴,相当于模拟设计专家个体的推理活动,属于简单系统;而人机智能化设计系统涉及多领域多学科知识范畴,是模拟和协助人类专家群体的推理决策活动,是人机复杂系统。

4) 从知识模型看,设计型专家系统只是围绕具体产品设计模型或针对设计过程某一特定环节(如有限元分析)的模型进行符号推理;而人机智能化设计系统则要考虑整个设计过程的模型、设计专家思维、推理和决策的模型(认知模型)以及设计对象(产品)的模型。

由此可见,人机智能化设计系统是针对大规模复杂产品设计的软件系统,它是面向集成的决策自动化,是高级的设计自动化。

智能设计作为计算机化的设计智能,乃是 CAD 的一个重要组成部分,它在 CAD 发展过程中有不同的表现形式。传统 CAD 系统中并无真正的智能成分,这一阶段的 CAD 系统虽然依托人类专家的设计智能,但作为计算机化的设计智能并不存在,智能设计在其中的作用也就无从谈起。而在 ICAD 阶段,智能设计是以设计型专家系统的形式出现的,但它仅仅是为解决设计中某些困难问题的局部需要而产生的,只是智能设计的初级阶段。对于 I2CAD 阶段,智能设计的表现形式是人机智能化设计系统,它顺应了市场对制造业的柔性、多样化、低成本、高质量、迅速响应能力的要求。作为 CIMS 大规模集成环境下的一个子系统,人机智能化设计系统乃是智能设计的高级阶段。上面讨论的有关设计技术和智能设计的若干概念及其对应关系可归纳为表 38.2-1。

表 38.2-1 设计技术及其说明

设计技术	代表形式	智能部分的承担者	说 明
传统设计技术	人工设计/传统 CAD	人类专家	
现代设计技术	ICAD	设计型专家系统	智能设计的初期阶段
先进设计技术	I2CAD	人机智能化设计系统	智能设计的高级阶段

#### (2) 智能设计技术

设计的本质是创造和革新。基于对设计本质的这

种认识,根据设计活动中创造性的大小,可将设计分为三类:常规设计(routine design)、革新设计(innovative

design) 和创新设计 (creative design)。显然, 革新设计是作为常规设计与创新设计的中介形式来界定的。所谓常规设计是指以成熟技术结构为基础, 运用常规方法来进行的产品设计, 它在工业生产中大量存在, 并且是一种经常性的工作。为了满足市场需求, 提高产品的竞争能力, 就需要改进老产品, 研制新品种, 降低生产材料、能源的消耗, 改进生产加工工艺等。在这种情况下, 就需要在设计中采用新的技术手段、技术原理和非常规方法, 即需要进行创造性设计。这里所说的创造性设计是创新设计和革新设计的统称。创新设计旨在提供具有社会价值的、新颖而独特成果的设计。它是设计探索中最富有挑战性的领域, 通常没有现成的设计规划, 有时甚至没有类似的已有设计

作为借鉴, 完全凭设计者去“无中生有”。革新设计是指为增加原有产品的功能、适用范围, 提高它的性能或改进其结构、尺寸或外形的变型设计, 因此也可称为是改进设计。这项任务实际上也包含了部分创造性内容, 但与“无中生有”相比, 它属于“举一反三”。

设计行为是思维活动的反映, 因而与人的思维密切相关。著名学者钱学森先生将人的思维划分为逻辑思维、形象思维和灵感思维三种形式, 并且指出实际上人的每个思维活动过程都不会是单纯的一种思维在起作用。三种思维形式的特点可归纳为表 38-2-2, 从中可知它们在创造性方面有不同的表现: 灵感思维最强, 形象思维次之, 逻辑思维最次。

表 38-2-2 思维的基本特点

思维形式	载体特点	特征
逻辑思维	一些抽象的概念、理论和数字等	抽象性、逻辑性、规律性、严密性, 思维过程是一维性
形象思维	形象, 如语言、图形、符号等	形象性、概括性、创造性、运动性, 思维过程是二维性
灵感思维	既可是抽象的概念等, 又可是形象	突发性、偶然性、独创性、模糊性, 思维过程是三维性

常规设计主要是通过逻辑思维实现的。创新设计通常是指采用发散而不是聚合的思维过程的设计, 这就使得形象思维乃至灵感思维在创新设计中显得更为关键和重要。

智能设计发展的不同阶段, 解决的主要问题也就不同。设计型专家系统解决的主要问题是模式设计, 方案设计可作为其典型代表, 它基本属于常规设计范畴, 但也包含一些革新设计的问题。与设计型专家系统不同, 人机智能化设计系统要解决的主要问题是创造性设计, 包括创新设计和革新设计。这是因为在 CIMS 这样的大规模知识集成环境中, 设计活动涉及了多领域、多学科的知识, 其影响因素错综复杂。当前颇为引人注目的并行工程 (concurrent engineering) 和并行设计 (concurrent design) 就鲜明地反映出面向集成的设计这一特点。CIMS 环境对设计活动的柔性提出了更高要求, 很难抽象提炼出有限的稳态模式。换言之, 即使存在设计模式的话, 设计模式也是千变万化, 几乎难以穷尽, 这样的设计活动必定更多地带有创造性色彩。

根据前面关于设计思维的论述, 设计型专家系统主要模拟的是人类专家的逻辑思维。人机智能化设计系统除了逻辑思维外, 主要模拟人类专家的形象思维, 甚至包括某些灵感思维。

## 3 智能设计体系和知识表达

### 3.1 智能设计体系

典型的设计过程是以设计师为主导完成的知识循环“迭代”过程, 可表示为“初始设计——评价——再

设计”, 即设计师根据实际要求, 先进行概念构思, 制定出初步的设计方案; 其次, 利用各种技术 (如有限元、优化设计等分析方法) 对方案进行评价和计算, 实现详细的具体设计; 最后, 对结果进行评价。当达到要求时, 设计完成; 当要求未达到时, 修改设计方案, 再进行第二轮的设计, 这样循环往复, 直到满足要求为止。

智能设计 (Intelligent Design—ID) 的目的是利用计算机全部或部分辅助代替设计师从事以上的整个设计过程, 在计算机上模拟或再现设计师的创造性设计过程。人工智能系统与一般计算机应用系统不同, 一般计算机系统处理的对象是数据, 而人工智能系统处理的对象可以是数据, 也可以是信息, 更重要的是处理各种知识, 使系统具有思维和推理能力。以往的研究集中在传统的数值计算和基于符号知识的推理的基础上进行, 进一步的研究迫切需要从更广泛的智能行为规律及内在运行机制进行探讨。

#### 3.1.1 智能设计的抽象层次模型

从问题描述的角度分析, 任何复杂系统都有必要抽象出统一的表达模型, 通过抽象可以把复杂的问题进行分层分类, 然后采用相应的处理方法。简言之, 复杂系统是由简单系统复合而成。以具有代表意义的复杂系统计算机网络为例, 计算机网络由各个节点 (节点处理机) 构成, 要实现节点与节点的通信而不造成系统的紊乱, 在计算机网络中引入了协议这一术语, 协议是为实现节点与节点间的同步与协调而作出的约定。著名的 ISO/OSI 参考模型 (七层协议) 为网络通信奠定了坚实的基础。用户可以在每层上进行通信。低层次上

的通信,用户考虑问题复杂些;高层上的通信,用户使用起来更方便,更简单。对等层上是协议,相邻层上有接口,其下一层为上一层提供服务,通过这种层次关系,构成了一个复杂而运行可靠的通信网络。

参考 ISO/OSI 模型,总结归纳智能设计自身的特点,提出如图 38.2-1 的智能设计抽象层次模型。图的左边层次体现了智能设计过程中层与层之间的相互关联,上一层以下一层为基础,下一层为上一层提供支持与服务,同时可以看出,每一层有其自己的任务,正是这样的分层与分类,才构成复杂系统设计的统一整体。图的右边体现了抽象层次模型在具体应用时所承担的任务,同样也呈现出如左边一样的特性。建立智能设计的抽象层次模型,是智能设计系统集成求解的基础。

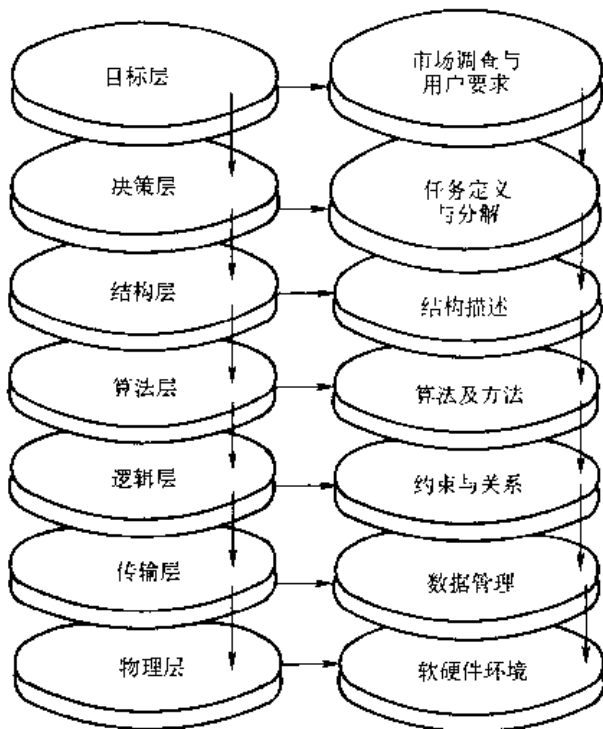


图 38.2-1 智能设计抽象层次模型

目标层为智能设计要达到的总目标,声明系统要达到的要求,往往与市场的需求,用户的要求相关联。

决策层把要实现的总目标分解成子目标,并采用相应的求解方法与策略,表现为任务的分解与进一步的决策。如智能设计中包含方案设计与布置设计,针对不同的设计要求,确定采用什么样的知识表达方法与求解策略。

结构层提供问题组织与表达的方法。结构层的合理确定,是保证系统统一和完整的先决条件。如目前广泛采用的面向对象的组织方式,可以为问题的描述提供有力的支持,结构层是实现集成的基础。

算法层是概念设计中最关键的一层,为决策层提

供强有力的支持工具。算法层包含所有可用的算法与方法。知识工程中的专家系统技术与基于实例的推理技术以及计算智能的人工神经网络、遗传算法都可以为决策层提供支持,是求解问题的关键所在。

逻辑层为算法层的协调、协作提供保障,逻辑层通过关系与约束把算法层沟通起来,使系统融合为一体。

传输层保证信息的交换,数据的管理,是以上各层信息交流的平台。

物理层提供系统运行的硬件环境,包括信息的存储以及与其他外部设备的连通。

### 3.1.2 设计知识的结构体系

工程设计属于复杂系统设计范畴,其特点是反复试验,不断摸索。从人类的思维形式角度来看,包含两种不同的方面,即抽象思维与形象思维。抽象思维是以抽象的概念和推论为形式的思维形式,概念是反映事物或现象的属性或本质的思维形式,掌握概念是进行抽象思维、从事科学创新活动的最基本的手段。形象思维是以形象化的“意象”为形式的思维形式,形象思维是理性认识,不是感性认识。“意象”是对同类事物形象的一般特征的反映,形象思维表现为人类思维的形象化与图视化,运用形象思维可以激发人们的想象力和联想、类比能力。

按照在思维过程是否严格遵守逻辑规则,可以将思维区分为逻辑思维和非逻辑思维两种方式。逻辑思维是严格遵循逻辑规则进行的一种思维方式,逻辑规则是人们在总结思维活动经验和规律的基础上概括出来的。逻辑思维以抽象的概念作为其思维元素,操作方式主要是分析和综合、归纳和演绎。非逻辑思维不严格遵守逻辑规则,表现为更具灵活性的自由思维,往往突破常规,具有鲜明的新奇性。非逻辑思维的基本形式是联想、想象、直觉和灵感。

从人类思维发展的角度来看,人类思维分为简单思维与复杂思维。简单思维与复杂思维最根本的区别在主体拥有知识的多少和主体对客体的认识程度。人类通过劳动和学习,在前人的基础上积累知识。随着知识的不断积累,知识的形式也呈现出多种多样;理论知识和实践知识。理论知识和实践知识体现了人类知识的不同层次关系,理论知识是实践知识的抽象化与升华,是具有抽象性、系统性和普遍指导意义的知识,是来自于实践知识而又与实践知识有本质区别的。实践知识是人类通过生产劳动而获得的知识,虽然还不具有抽象性和系统性,但具有实用性,是在理论知识的指导下而产生的知识。理论知识和实践知识互相促进,相互转化,螺旋式地向前发展。对设计知识的认识和处理是再现人类思维规律的基础。从认识论的角度分析,人



类知识可分为三类：

- 1) 过程性知识 (procedural knowledge)。
- 2) 叙述性知识 (declarative knowledge)。
- 3) 潜意识 (tacit knowledge)。

过程性知识是对客观事物的精确描述，可以用准确的数学模型来表达。例如：传统的优化设计，首先对问题进行描述，确定设计变量，约束条件及目标函数，在此基础上选用适当的优化求解方法，通过计算机的数字迭代，求解出满足要求的设计变量值。涉及到数学模型的建立，求解速度和收敛性的分析。采用过程性知识进行问题求解的前提是待求问题的性态要求结构优良，易于收敛。

叙述性知识是指对客观事物的描述能够用语言文字来表达，既可方便地将人类知识以明确规范化的语言表达出来，也便于计算机的实现。这种问题不能用严密的数学模型来刻画。叙述性知识大多表现为人类专家经验知识的归纳，以符号的形式存在。

潜意识是指客观事物不能或难于用明确规范化的语言表达出来，即使专家本身也很难说出他们的理由，具有很强的跳跃性和非结构性。而往往这种知识是创造性设计的关键。潜意识表现为人类专家经验知识量积累到一定程度以后的一个质的飞跃，用这些经验（比如以往设计成功的设计范例）通过联想“想当然”地作出快速的决策。

设计师在设计时，采用的知识并不是单一的。由于问题的复杂性，决定了知识的异构性。过程性知识、叙述性知识及潜意识为异构知识的抽象形式，更具体化的形式可以概括为：过程知识、符号知识、实例知识和样本知识。通过以上分析，作出如下异构知识的定义

**定义** 异构知识：工程设计中不同层次、不同表现形式的异性知识构成了异构知识 IK (isomeric knowledge, 简称 IK)，抽象描述为：

$$IK = (PK, SK_1, CK, SK_2)$$

- 式中 PK——Procedural Knowledge，即过程知识；  
 SK<sub>1</sub>——Symbolic Knowledge，即符号知识；  
 CK——Case Knowledge，即实例知识；  
 SK<sub>2</sub>——Sample Knowledge，即样本知识。

过程知识、符号知识、实例知识和样本知识构成了异构知识体系。在这异构知识体系中，不同层次、不同形式的知识相辅相成，互为补充。

### 3.1.3 智能设计的集成求解策略

建立智能设计抽象层次模型，目的是实现智能设计系统集成求解。基于抽象层次模型的系统集成的关键是在人工智能领域探索有效的求解途径。人工智能目前主要分两大流派：符号主义流派和联接主义流派。符号主义流派以专家系统 (expert system, 简称 ES) 和基于实例推理 (case-based reasoning, 简称 CBR) 为代表——统称为知识工程 (knowledge engineering, 简称 KE)。联接主义流派以人工神经网络 (artificial neural network, 简称 ANN) 和遗传算法 (genetic algorithms, 简称 GA) 为代表——统称计算智能 (computational intelligence, 简称 CI)。图 38.2-2 表征了这种关系。

图 38.2-3 描述知识工程 (KE) 与计算智能 (CI) 相结合的异构知识求解模型，用 ES、CBR 及 ANN 求解异构知识。智能设计 (ID) 为问题求解的核心（核心圆）；第一环描述了人类的思维模式，即抽象思维、形象思维、逻辑思维和逻辑思维；第二环描述了思维处理的异构知识，即过程知识、符号知识、实例知识和样本知识；第三环描述了相应的求解方法（途径）。表 38.2-3 为四种人工智能方法的比较结果。由此看来，四种人工智能的结合，具有强大的求解能力。

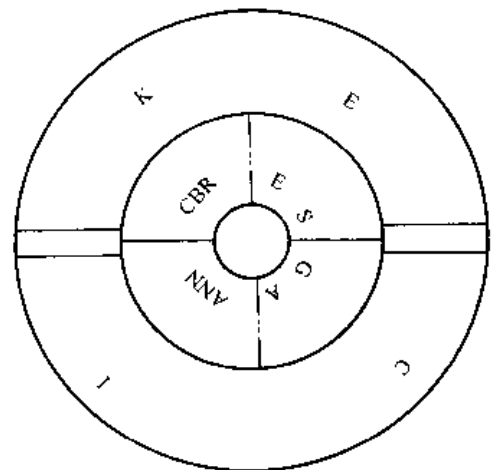


图 38.2-2 知识工程与计算智能分类

表 38.2-3 四种人工智能方法的比较

方法	优化能力	思维方式	学习能力	知识的可操作性	解释功能	知识形式	非线性能力
ES	较弱	抽象思维	较差	有	强	过程、符号	弱
CBR	有一些	类比思维	较强	有一些	有一些	实例	有
ANN	较强	联想思维	强	无	无	样本	强
GA	强	仿自然	有	无	无	多种知识	强



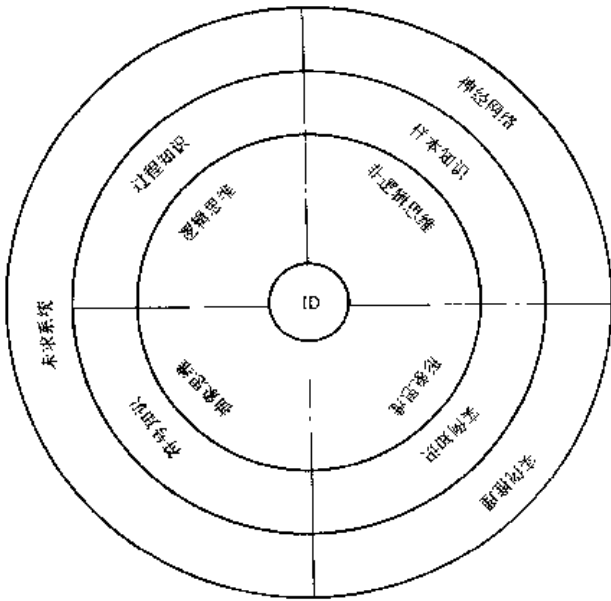


图 38.2-3 知识工程与计算智能相结合的知识求解模型

### 3.1.4 智能设计集成求解策略工程应用

智能设计体系结构的研究是为了解决复杂工程设计问题。根据工程设计基本特征,需要解决以下困难问题:①异构知识的处理;②多方案设计;③再设计;④设计效率的提高,介绍如图 38.2-4 所示的智能设计集成求解策略。

对于基于符号知识的推理求解来说,初始设计通过专家知识的推理得到初步方案,再进一步分析推理结果,然后评价其结果是否满意。如果结果满意,输出结果;如果结果不满意,修改相关参数,重新确定新的方案,重复以上步骤直到结果满意为止。基于符号知识的推理求解符号性知识和过程性知识,属于逻辑思维。由于工程问题的复杂性,基于符号知识推理技术在多方案的产生和再设计问题上非常困难,基因算法为多方案的产生提供了有效的机制,而约束满足方法则为基于符号知识推理提供了有效的再设计手段。

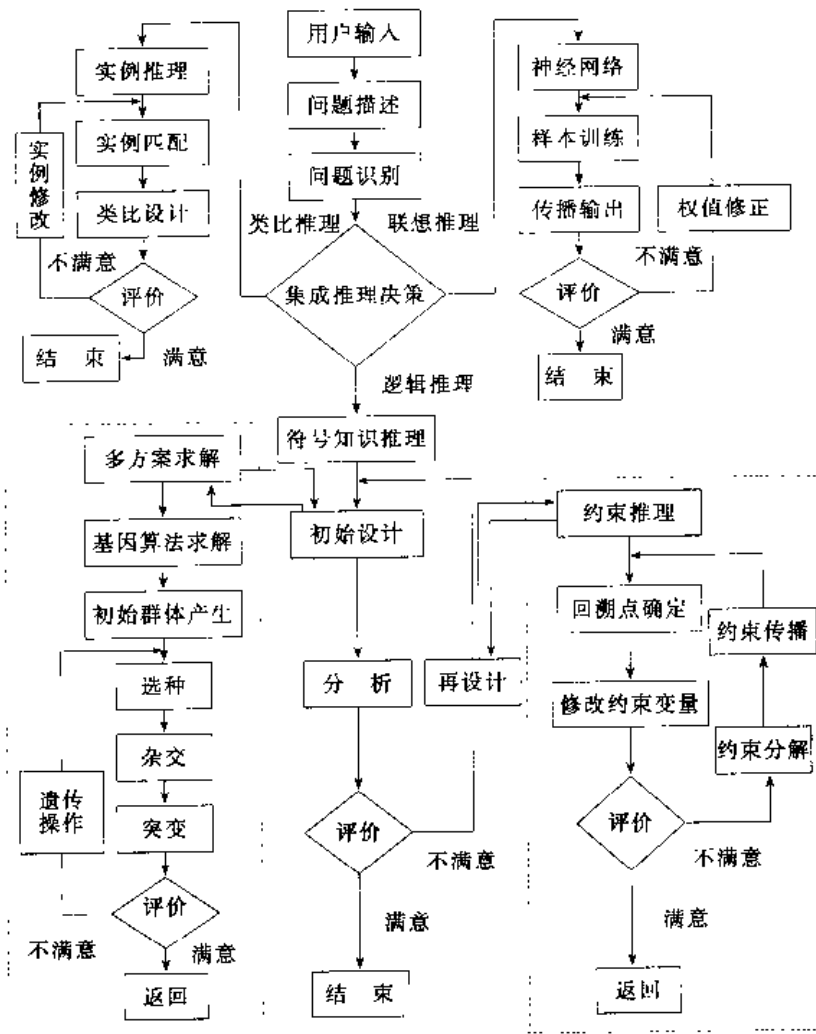


图 38.2-4 智能设计集成求解策略

对于基于实例推理求解来说,初始设计是提取相关实例,对相关实例进行类比设计,再通过实例的评价,确定是否采用该实例,或进一步修改实例以满足设计要求。基于实例推理求解实例知识,属于类比思维。

对于人工神经网络求解来说,初始设计是在样本训练的基础上,通过输入值的传播产生候选解,对候选解进行评价,若不满意输出结果,可重新调整网络数值,或增加样本,或提炼样本,改进误差,直到输出结果满意为止。人工神经网络学习处理样本知识,属于直觉思维(“潜意识”)。

对于采用基因算法求解来说,初始设计是通过随机产生个体,再由个体的选择、重组、杂交、突变,然后施用进化压力,使个体往优良的方向发展,如果得到的个体最优则输出,否则进一步通过遗传操作修改个体,直到使个体满意为止。基因算法为基于符号知识推理快速提供初始多方案设计。

### 3.2 智能设计的知识表达

产品概念设计是一个复杂的、不完全确定的、创造性的过程,它是最终实现智能化设计的关键环节,而设计过程合理有效的知识表达是智能设计的基础。概念设计过程是由分析用户需求到生成概念产品的过程,它实际上是一连串相连的问题求解活动。“每一种问题求解方法都需要对某种解答的搜索,不过,在搜索解答过程开始之前,必须先用某种方法来表达问题。任何比较复杂的求解技术都离不开两方面的内容——表达与搜索”。而概念设计问题求解过程的知识表达的优劣,对求解结果及求解工作量的影响很大,它是问题求解的第一步工作。这里讨论的知识是指与在特定的专门领域中进行问题求解过程有直接关系的知识。

#### (1) 概念设计问题规约

人工智能原理中的问题归约法(Problem Reduction)是一种问题描述与求解方法:把一个复杂的原始问题分解为若干个较为简单的子问题,每个子问题又可继续分解为若干个更为简单的子问题,重复此过程,直到不需要再分解或者不能再分解为止(这种不能再分解或变换,而且直接可解的子问题称之为本原问题)。然后对每一个子问题(本原问题)分别进行求解。最后把各个子问题的解复合起来就得到了原始问题的解。

产品的概念设计的原始问题可以通过一系列分解变换归约为一个本原问题的集合,然后经过对每一个本原问题分别进行求解,最后把各个本原问题的解综合起来的过程,就可以得到概念设计原始问题的解,如图 38.2-5 所示。在这里,一个抽象化的设计问题 ADP 可以被看作为原始问题(用方框表示),功能元  $E_{Fi}$  可

以被看作为本原问题(用方框表示),而最后的最优原理方案 CP(用方框表示)则是原始问题 ADP 的解。

#### (2) 概念设计问题的与/或树表达

1) ‘与/或’树的定义 ‘与/或’树(And/Or Tree)是用于表示问题及其求解过程的一种形式化方法。例如:如把问题  $P$  分解成 3 个子问题  $P_1$ 、 $P_2$ 、 $P_3$ ,如图 38.2-6a 所示。只有 3 个子问题都可解时,问题  $P$  才可解,称  $P_1$ 、 $P_2$ 、 $P_3$  之间存在‘与’关系(用圆弧线连接),并称节点  $P$  为‘与’节点。 $P$ 、 $P_1$ 、 $P_2$ 、 $P_3$  所构成的图成为‘与’树,用符号表示为: $P \rightarrow \{P_1 \wedge P_2 \wedge P_3\}$ ,  $\rightarrow$  表示等价。

如把问题  $P$  变换成 3 个子问题  $P_1$ 、 $P_2$ 、 $P_3$ ,如图 38.2-6b 所示。3 个子问题中只要有一个可解时,问题  $P$  就可解,称  $P_1$ 、 $P_2$ 、 $P_3$  之间存在‘或’关系,并称节点  $P$  为‘或’节点。 $P$ 、 $P_1$ 、 $P_2$ 、 $P_3$  所构成的图成为‘或’树,用符号表示为: $P \rightarrow \{P_1 \vee P_2 \vee P_3\}$ 。

把上述两种方法结合起来,就构成了‘与/或’树,如图 38.2-7 所示。原始问题对应的节点称为初始节点;子问题对应的节点称为子节点;本原问题对应的节点称为终止节点。

2) 问题的‘与/或’树表达 按照上述规定的‘与/或’树表达方法,以及图 38.2-5 所建立的问题归约模型,可以得到产品概念设计问题的‘与/或’表达树,如图 38.2-8 所示。图中一个抽象的产品概念设计问题(ADP)作为原始问题对应着‘与/或’树的初始节点(用方框表示);问题分解后所得到的功能元( $E_{Fi}$ )作为本原问题对应着终止节点(用方框表示);原理解答对应着原理解节点( $PS_i$ );优化的概念产品方案(CP)作为最终解答对应着最后的节点(用方框表示);其余的圆形节点作为子节点。当其各边用一条圆弧线连接时,称之为‘与’节点,否则为‘或’节点。

#### (3) 原始问题解的判定

1) 可解节点判定 在图 38.2-8 所示的概念设计问题‘与/或’树中,满足下列条件之一的节点,称之为可解节点。

① 它是一个原理解节点( $PS_i$ )。

② 它是一个‘与’节点,且其子节点全部是可解节点。

③ 它是一个‘或’节点,且其子节点中至少有一个是可解节点。

上述三个条件都不满足的节点为不可解节点。

2) 原始问题解的判定 在图 38.2-8 所示的产品概念设计‘与/或’树中,原始问题(初始节点)有组合原理解集合的条件为:如果‘与/或’树中的所有节点都是可解节点,则存在组合原理解集合:

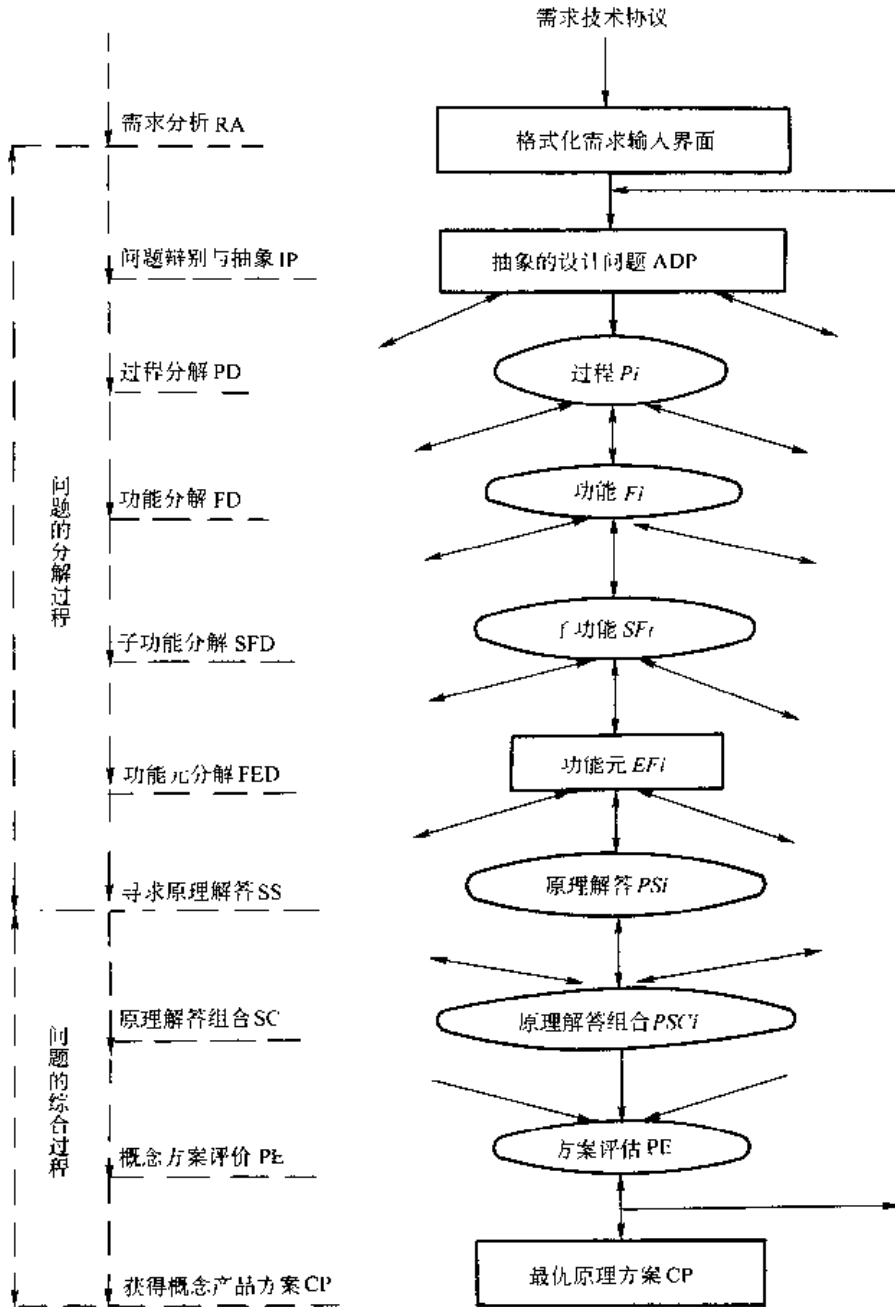


图 38.2-5 概念设计问题归约模型

$$PSC = \{PSC_k \quad k \in T\}$$

式中  $T = \{1, 2, 3 \dots\}$ 。

如果‘与/或’树存在不可解节点，则不存在组合原理解集合。此时将需要重新进行局部的、甚至是全局的问题归约或借助于自动推理过程以外的、人的干预和帮助。

(4) 原始问题解的表达

1) 符号定义

① 连接词：

$\vee$ ：称为‘析取’。它表示被它连接的两个命题具有‘或者’关系。

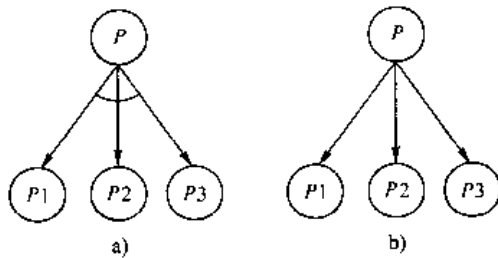


图 38.2-6 ‘与’树和‘或’树

$\wedge$ : 称为‘合取’。它表示被它连接的两个命题具有‘与’关系。

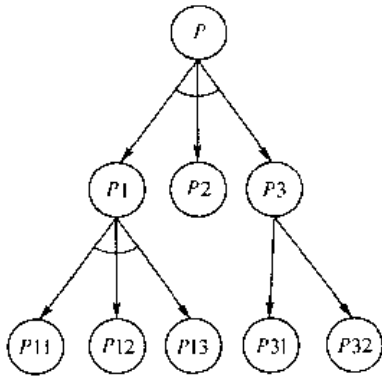


图 38.2-7 ‘与/或’树

$\Rightarrow$ : 称为‘条件’。 $P \Rightarrow Q$  表示‘如果 P, 则 Q’, 其中的 P 称为条件的前件, Q 称为条件的后件。

② 量词:

$(\forall x)$ : 称为‘全称量词’。它表示‘对于个体域中的所有 x (或任意一个个体 x)’。

$(\exists x)$ : 称为‘存在量词’。它表示‘对于个体域中存在个体 x’。

2) 原始问题解的表达 在图 38.2-8 中, 如果每个终止节点  $EF_j$  (本原问题) 都存在若干原理解  $PS_{ij}$ , 则原始问题存在组合原理解集合。

① 任意组合原理解的谓词逻辑表达为:

$$(\forall j \in S) [\exists PS_{ij} \in EF_j] \rightarrow$$

$$[(\exists PS_{i1} \in EF_1) \wedge (\exists PS_{i2} \in EF_2) \wedge (\exists PS_{i3} \in EF_3) \wedge \dots \wedge (\exists PS_{iq} \in EF_q)]$$

$$= (\exists k \in T) PSC^k$$

式中  $S = \{1, 2, 3, \dots, q\}$

$T = \{1, 2, 3, \dots\}$

$i_1 \in T, i_2 \in T, \dots, i_q \in T$

② 全部组合原理解集合的谓词逻辑表达为:

$$(\forall j \in S) [\forall PS_{ij} \in EF_j]$$

$$\Rightarrow \bigwedge_{j=1}^q [\bigvee_{i=1}^n PS_{ij} \in EF_j]$$

$$= PSC = \{PSC^k \mid k \in T\}$$

式中  $S = \{1, 2, 3, \dots, q\}$

$T = \{1, 2, 3, \dots\}$

$i_1 \in T, i_2 \in T, \dots, i_q \in T$

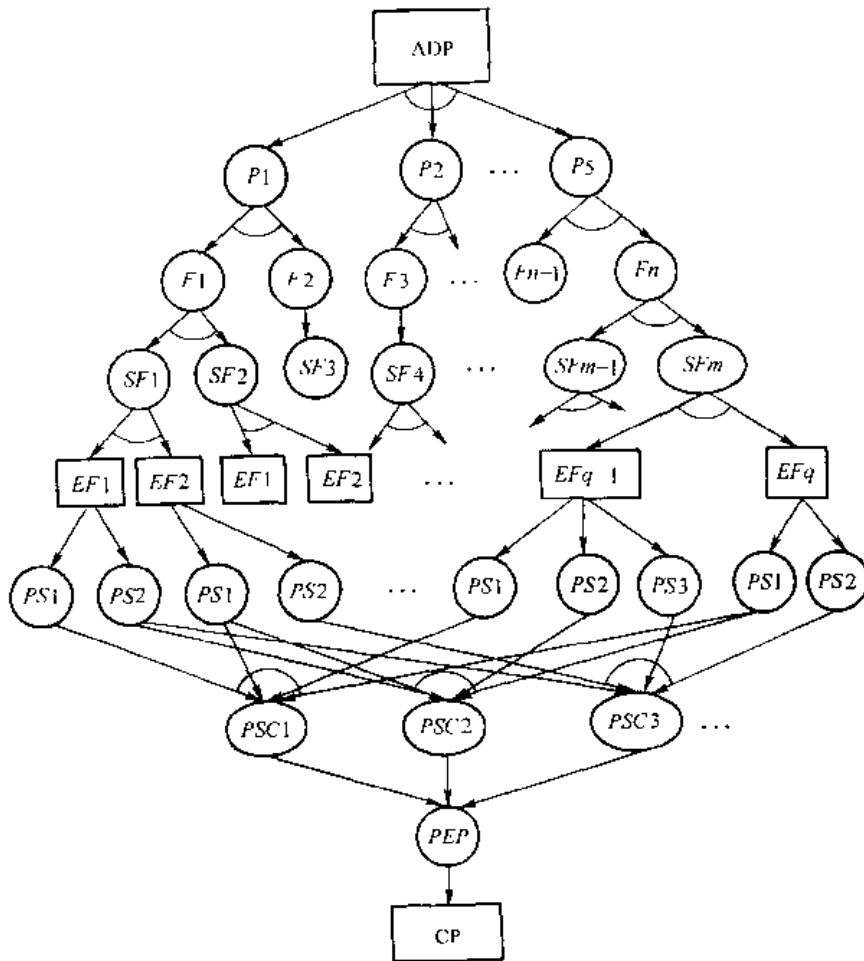


图 38.2.8 概念设计问题 ‘与/或’ 树

### 3.3 智能设计的基因模型表达

#### 3.3.1 知识模型

机械智能 CAD 系统的核心任务是关于设计的知识问题,即知识的表示、获取和应用问题。智能系统在知识学习和运用上的突破,首先必须是知识模型的突破,从而实现知识体系的突破,进而实现系统设计能力的突破,即由常规设计向联想设计、进化设计的突破。

智能设计中,提出的知识模型与智能设计能力层

次的对应关系如表 38.2-4 所示。

所谓知识模型,是智能设计中基于领域知识的对应某一设计层次的设计模型,该设计模型是与各设计层次采用的技术方法相一致,并按相应技术方法的特点对领域知识进行组织、表示、获取和应用,其特点如下:

1) 对应的技术方法不同,则知识模型的组织表现形式、获取应用方法不同。例如在关于发动机活塞环组的低功耗设计中,常规设计的符号模型表示规则为:

IF (功耗 > 设计阈值) THEN (减小第一道活塞环宽)

表 38.2-4 设计知识模型与智能设计能力层次

智能设计能力层次 \ 特点	知识模型	知识库	设计规则类型	性能
常规设计	符号模型、CAD 模型	显形知识规则库、符号模拟等	关于单个参数的设计经验	适用范围有限(良好问题)
联想设计	神经网络模型	网络参数库	反映多个设计参数影响及影响大小排序	推理迅速,但训练过程依赖于网络初值和结构
进化设计	基因模型	基因块库	反映多个影响参数组合并行影响	设计创新,是反馈式自适应过程,但设计解获取时间较长

联想设计则用人工神经网络模型来建立功耗与诸影响因素的映射关系,如图 38.2-9 所示。

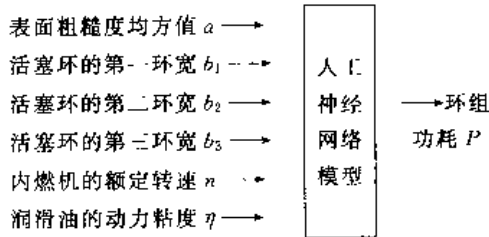


图 38.2-9 活塞环组功耗的网络模型

网络由样本数据训练完成后,关于诸参数对功耗的影响的知识隐含在网络的结构和权值中,成为相应知识的神经网络模型。

2) 不同的知识模型适合不同的设计问题。符号模型适合于表现再设计规则,神经网络模型适合于拟合复杂的非线性函数关系,而基因模型则适宜在设计过程中自适应地得到方案的优选结果。例如采用基因模型,可以在一定机型、一定设计要求下得到优良设计方案中参数的组合关系:

$$b_1 : b_2 : b_3 \text{ 的优选值大约为 } 1.0 : 1.2 : 1.1$$

其中  $b_1$ 、 $b_2$ 、 $b_3$  分别为第一、二、三道活塞环宽。可见,基因模型不仅可以并行调整设计参数,提高效率,而且可以实现设计方案组合的优化创新。

#### 3.3.2 基因模型

基因模型是指采用遗传算法的编码规则表示设计

方案和设计知识。由此设计过程转换为基因样本种群的进化过程,如表 38.2-5 所示。

表 38.2-5 遗传算法和设计领域的对应简表

进化计算	设计领域
一个基因串	一个设计方案
基因串的群体	设计方案的群体
适应函数	设计方案评价标准
进化的基因串	设计优良的方案
基因块	设计优良解的参数组合特征

对工程设计来说,进化计算、基因模型的研究可以归结为三个方向:

#### (1) 设计参数的优化设计与调整

工程设计中,尤其大型复杂机械系统的设计往往涉及参数类型多样(连续变量、离散变量、整型变量等)、设计空间不连续、梯度信息匮乏、设计约束多样而且相互冲突等问题,使传统的优化、搜索技术已经不能满足要求,而进化计算恰恰可以用于求得优化解或可行解。对传统优化方法的补充是进化设计的主要内容,麻省理工学院的 A. Thornton 博士将设计问题视为设计约束下的参数优化问题,将设计约束转变成为设计目的的“罚函数”,试图在设计与优化之间达到统一。

而实际的设计过程中,更关心的是设计方案的确定、设计满足解的寻求,没有必要也不可能求出设计最优解,如美国通用电器公司和 Rensselaer 综合学院的学者将遗传算法应用于喷气发电机的涡轮设计之中,

涉及至少 100 个设计变量,每个变量取不同范围的值,搜索空间的点数不少于  $10^{30}$  个,评价 1 个点(对应一个设计方案)需要运行的发电机模拟程序在一般工作站需要 30s,该计划将花费 5 年或更长的时间,预计耗资 20 亿美元,即使这样,仍需要专家系统技术的支持,把专家系统生成的初始设计方案作为遗传算法的起点来提高效率。

### (2) 设计知识的获取与应用

首先在经验的提取方面,由于良好的设计样本具有高繁殖率,所以可以进行优良样本信息的提取,参考文献 [10] 用遗传算法提取相似事例对应的基因编码,作为设计规则应用于当前设计中,指导进化过程。其次,在设计知识的应用方面,基于遗传算法的搜索以产生创新解为特色,以搜索的空间和时间作为代价,领域知识的运用可以指导解的搜索范围,提高效率。A. Thornton 博士提出一种“屏蔽算子”(Masked crossover),包含了约束空间的知识,只有通过这一算子的基因样本才能存活,研究表明,其效率远远大于传统的单点或两点交叉算子。

### (3) 设计方案的创新

悉尼大学以 M. A. Rosenman 和 J. Gero 为代表的非常规设计研究小组,将进化设计应用到建筑设计问题中,实现轮廓布置方案的组合创新,他们认为复杂的布置问题由简单的建筑元件采用简单的派生规则进化

形成,并采用最简单外形的建筑元件(正方形)作为研究对象,元件的派生规则只有简单的四种,即右派生、下派生、左派生和上派生,用二进制分别表示为(00, 01, 10, 11),布置设计方案表现为派生规则的顺序组合,如基因串(011010111100)表示的布置方案如图 38.2-10 所示。

多个基因串表示的方案群体,由建筑专家决定样本的优劣,好的样本进行复制,交叉、变异后,形成设计种群的下一代(样本基因交叉、变异的过程是新的方案生成的过程),如此不断进化直到产生出满意解(内含优选解和创新解),这项研究提出了设计方案表述的新思路。遗憾的是,由于建筑布置几何直观性的特点,决定其进化适应函数,进化算子较难符合机械设计的特点。

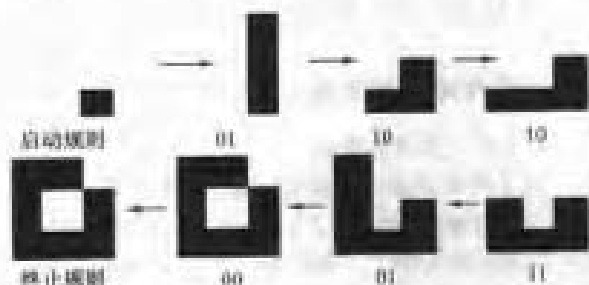


图 38.2-10 基因串(011010111100)表示的布置方案

## 第3章 进化设计技术与方法

进化设计技术与方法是以进化算法 (Evolution Algorithms, 简称 EA) 为基础的。进化算法是与进化计算相关的算法的统称。进化算法主要包括遗传算法 (Genetic Algorithms)、遗传规划 (Genetic Programming)、进化策略 (Evolutionary Strategy)、进化规划 (Evolutionary Programming) 和模拟退火算法 (Simulated Annealing Algorithms) 等, 而以遗传算法最具有代表性。本章在简述进化设计技术基础后, 重点地介绍几种进化设计方法。

### 1 进化设计技术基础

#### 1.1 遗传算法的概貌

##### (1) 什么是遗传算法

为了简单而通俗地说明问题, 首先在图 38.3-1 中定义一个变量  $x$  的函数  $f(x)$ , 且  $f(x)$  在区间  $[x_1, x_2]$  内存在最大值  $f_{\max}$  或最小值  $f_{\min}$ 。GA 的功能之一就是快速求解类似这类问题的最大值  $f_{\max}$  或最小值  $f_{\min}$ 。一般地讲, GA 以模拟生物进化过程为基本原理, 是适用于所有最优化搜索问题的方法。GA 在计算机中设定假想生物集团, 其中适应所处环境的生物个体, 其生存概率也相应高。利用生物集团中个体的适者生存、劣者淘汰的仿真过程, 实现基因和生物集团的进化。由生物集团的进化思想, 用计算机仿真解决工程课题, 需要编制 GA 的程序。但是, GA 的编程存在以下特点:

- 1) 程序无详细的模式。
- 2) 各种规则和参数设定的不确定性。

这些特点, 从一个侧面来讲是 GA 的缺点, 从另一个侧面来讲也决定了 GA 对解决各类问题具有柔软的适应性, 可广泛地适用于各类最优化问题。追溯 GA 的发展历史, 20 多年前在美国开始对其研究, 至今已在工程应用各领域产生了巨大的影响。今后其研究必将具有广阔的前景。

##### (2) 遗传算法的最优搜索

GA 适用于无符号整数变量空间的最优化搜索问题。设图 38.3-1 的变量  $J$  为整数 ( $x \in [0, 255]$ ) 的前提下, 用简单的循环算法, 可顺序求出  $f(0), f(1), \dots, f(255)$  的值, 即进行 256 次计算可得到  $f_{\max}$  的值。

可是, 一般此类搜索问题的变量  $y$  的定义域是很

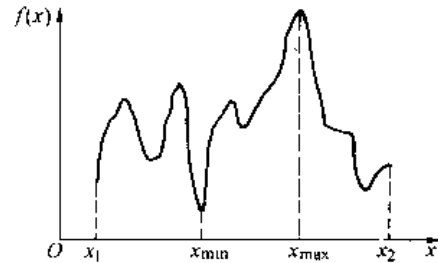


图 38.3-1 函数  $f(x)$  的最大值或最小值搜索问题

大的。例如, 变量  $y \in [y_1, y_2]$ , 则可将  $y$  的实数域  $[y_1, y_2]$  线性映射到无符号整数区域  $[0, 2^n]$ , 其中  $n$  为 1000 的话, 变量  $y$  的编码共  $2^{1000} \approx 10^{300}$  个。假如, 计算机每秒计算 1000 个搜索点, 一天中只能计算  $1000 \times 60 \times 60 \times 24 \approx 10^8$  个, 所以全部计算需要  $10^{100} \div 10^8 \approx 10^{252}$  天  $\approx 10^{286}$  年。  $10^8$  年是 1 亿年。可以想象对此类问题, 从原理上是可解的, 但实际上却是不可能得到解的, 通常把它叫做 NP 问题。从工程最优化搜索的角度来讲, 并不一定要求出其真正的最优解, 只要能求出其最优解附近的较优解, 在实用上已可满足。GA 在离散组合空间具有高效率搜索的特点, GA 的方法可以高效率地搜索设定的搜索空间, 得到满足实用的较优解, 不失为一种最优搜索的方法。

##### (3) 遗传算法的基本思路

GA 在搜索空间中的搜索并不是以单个搜索点顺序进行的, 而是使用若干个搜索点 (生物个体) 并行进行搜索的。一个搜索点, 作为一个带有遗传情报的假想生物个体, 即简单作为一个生物个体来对待。若干个生物个体即构成了生物集团。首先, 相对于各生物个体, 计算出其生物个体对所处环境的适应度。以图 38.3-1 所示为例, 可以把  $x$  作为个体, 把  $f(x)$  作为  $x$  所处环境的适应度来处理。然后, 淘汰适应度低的个体, 增殖适应度高的个体。如此进行世代交替仿真, 实现进化 (最优化) 计算。算法中相对于实际的增殖, GA 以基因型的交叉以及突然变异操作来进行。最后求出非常高的个体, 即  $\max(f(x))$  的  $x_{\max}$  值。

以上就是 GA 的基本思路。其具体的操作以图 38.3-2 来形象地说明。GA 的内部处理过程先作为一个未知的黑盒看待。相对给其一定数位 0-1 字符串的输入, 将会得到其评价。问题是输入怎样的 0-1 字符串, 才能得到好的评价。假如随意地输入若干个 0-1 字符串, 比较出其中比较高评价点所具有的字符串前缀

下,在考虑哪些部分会给大的评价点的同时,进行字符串的局部复制或变更,制作出新的字符串群。如此反复进行,使字符串群所具有的平均评价点升值。其过程大略如下:

1) 首先随机地产生字符串。

2) 参照高评价点的字符串群,通过部分复制和部分修正产生新的字符串群。在全体评价点低的情况时,对字符串进行大幅度的修正。

3) 在全体评价点高时,对字符串进行小幅度的修正,更详尽地决定字符串的细部。

GA 的操作就是实现对于个体的基因型进行这样简单基本操作的反复过程。

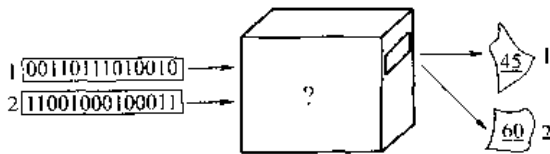


图 38.3-2 遗传算法的基本思路

### 1.2 单纯型遗传算法

单纯型遗传算法反映了 GA 的基本思路和操作方 法,对于理解 GA 是极其重要的。在本节里,概要地阐述单纯型遗传算法 (SGA) 的计算流程。

#### (1) 假想生物及其环境的设定

在进行假想生物集团的仿真之前,必须进行若干设定。

1) 设定个体的染色体和基因 首先,设定个体 (individual) 的染色体 (chromosome),即决策矢变量的编码字符串,也就是设定假想生物在进行生殖时,上世代个体把怎样的数据内容,以怎样的形式遗传给下世代的子孙个体。染色体如图 38.3-3 所示,一般由若干个基因 (gene) 构成。矢变量中的各分量则对应于各个基因,所处的位置为各基因位 (gene locus)。基因位可理解为表现染色体中各基因所在位置的一种坐标。

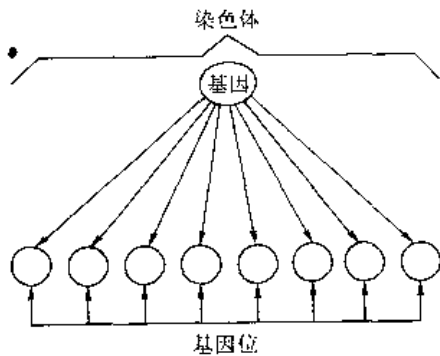


图 38.3-3 染色体、基因和基因位

各个体 (染色体) 的内部表现形式叫做基因型 (gene type),即矢变量的编码结构。表现基因型的形

式可以是任意的,但是一般使用 0 和 1 的排列方式。在这种方式下,每一基因位的基因用几位字符串来表示,要根据实际决策矢变量中各矢变分量的具体情况而定。例如,某矢变分量为—电路开关的闭合,其可能状态只有两种,用一位字符即可表示其基因。再例如,图 38.3-1 中变量  $x$  为整数 ( $x \in [0, 255]$ ),则用 8 位字符串作为其基因。再之,图 38.3-1 的最优化问题的矢变量只有一个分量,其基因型也就是其基因本身,用 8 位字符串表示。有关基因型的字符串长度,一般为固定长度形式,但随着世代交替的进行,让其基因型趋于复杂化,以满足进化的需要,也可以选择可变基因型长度的 GA 算法。

2) 表现型的设定 将基因型经过某种变换处理后的结构形式,叫做表现型 (phenotype),即决策矢变量的解码结构。在许多最优化问题中,表现型=基因型。有些情况下,随着世代交替,基因型变得复杂化和多样化,使得进化过程变得复杂和困难。为此,需要对基因型进行某种变换处理,用表现型来表示。表现型的设定方法没有固定的模式,必须根据问题的实际设定变换处理的方法。在图 38.3-1 中,用 8 位字符串表示整数决策变量  $x \in [0, 255]$  的表现型,即是表现型=基因型。

3) 设定适应度的计算方法 基因型、表现型设定后,应该设定表示各个体对环境的适应能力——适应度 (fitness) 的计算方法。适应度就是把搜索空间中的各搜索点作为各个体对待,相当于用个体所持的遗传信息来表现其所在空间位置的相应目标评价。图 38.3-1 中的最大值搜索问题,可用各个体的基因型所表示的变量  $x$  位置的  $f(x)$  值,作为适应度来处理。

适应度的计算方法也没有固定的格式,必须根据实际问题适应地设定。一般地讲,求解问题比较复杂,其适应度的计算方法也将会较复杂。再者,某个体的适应度,并非一定是同其他个体无关而进行简单计算得来的。有时也需要考虑同其他个体的关系而设定适当的计算方法。例如 GA 在人工生命 (artificial life-AL) 的应用中就必须考虑。总之,为了在假想生物进化中反映自然淘汰 (natural selection) 的原理,适应度从各个体生存的可能性角度,给出了评价个体、表现个体的一个定量尺度。

#### (2) 单纯型遗传算法的计算流程

在以上假想生物和环境的设定完成后,SGA 将服从图 38.3-4 所示的计算流程,使假想生物集团 (population) 进化。以下按图 38.3-4 的框图顺序给以简单说明。

1) 产生初始生物集团 GA 在搜索空间中设定若下个个体 (即搜索点),由这些个体组成生物集团。搜



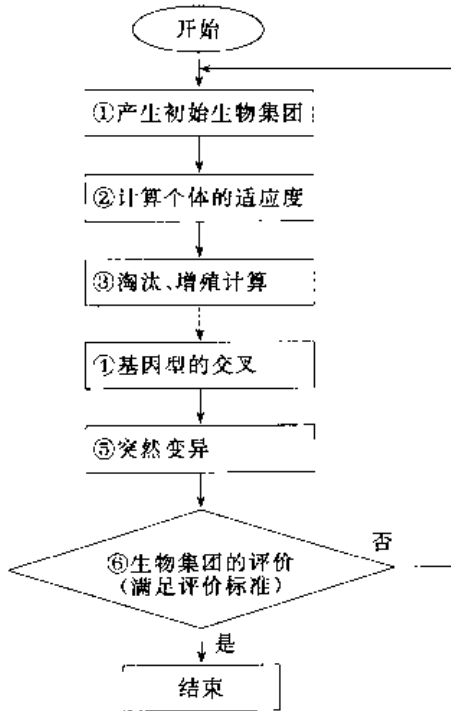


图 38.3-4 SGA 的处理流程

索开始时,问题的解完全是未知的,因此设定怎样的个体较好也完全是未知的。通常,初始生物集团用随机数随机发生。产生的个体总数为  $N$ , 个体用  $I_i (i=1, 2, \dots, N)$  表示。个体  $I_i$  的基因型用  $G_i$  表示。

对图 38.3-1 的最大值搜索问题,用下式随机地设定初始生物集团:

$$G_i = \text{rnd}(255) \quad (38.3-1)$$

式中  $\text{rnd}(n)$  ( $n$  为自然数) 是产生从 0 到  $n$  随机数的函数。

2) 计算个体的适应度 计算生物集团中个体  $I_i$  与环境的适应度  $F(I_i)$ 。

3) 淘汰、增殖计算 SGA 的生物集团淘汰 (selection) 和增殖 (multiplication) 处理,由简单的生殖 (reproduction) 处理构成。从现世代  $N$  个个体  $I_1 \sim I_N$  中,允许重复随机地选择  $N$  个个体,决定出下世代的  $N$  个个体。某个体  $I_i$  作为下世代个体,其被选中的概率为  $P(I_i)$ ,用下式计算:

$$P(I_i) = \frac{F(I_i)}{\frac{1}{N} \sum_{i=1}^N F(I_i)}$$

上式中,右边的分子是个体  $I_i$  的适应度,分母是现世代生物集团的平均适应度。此即为各个体在下世代生存的可能性与自身的适应度成比例。因此适应度高的个体,作为下世代个体被选中的概率也就越大。决定下世代个体的处理,可形象地用图 38.3-5 的轮盘赌 (roulette) 来说明。

在图 38.3-5 中,设定对应于各个体  $I_i (i=1, 2,$

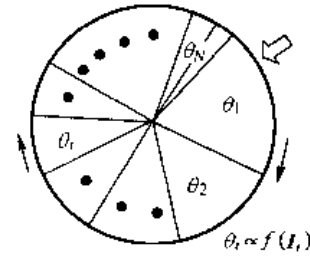


图 38.3-5 与适应度成比例的选择概率

$\dots, N)$  所占有的扇形角度  $\theta_i (i=1, 2, \dots, N)$  与  $f(I_i) (i=1, 2, \dots, N)$  成比例。SGA 中的淘汰和增殖操作,就是将轮盘赌随机地转动  $N$  次,每次箭头与轮盘位置相重合 (图 38.3-5),即与适应度成比例,概率地决定选中的个体。如此操作就等价于允许重复决定  $N$  个下世代个体的做法。

在这种选择方式中,适应度高的个体作为下世代个体被选中的可能性就大,即使相对于适应度低的个体,也存在着被选中为下世代个体的可能性。只选择适应度高的个体,固然其生物集团的收敛速度会快,但也易陷入局部最优解的误区。

如图 38.3-6 所示那样,假如现世代的个体有  $I_1 \sim I_5$ , 下世代只选择了  $I_1 \sim I_3$ , 适应度高的 3 个个体,可以直观地看出,因为下世代生物集团全部汇集在极大值  $P_1$  的周围,个体不能到达最大值  $P_2$  的可能性很大。相对于这样的问题,让生殖的下世代个体在适应度最低的 5 的附近也具有存在的可能的话,个体到达  $P_2$  的可能性也会大。也正因为如此,为了防止在淘汰和增殖操作中有可能出现生物集团失去多样性的问题,还将进行基因型的交叉和突然变异操作。

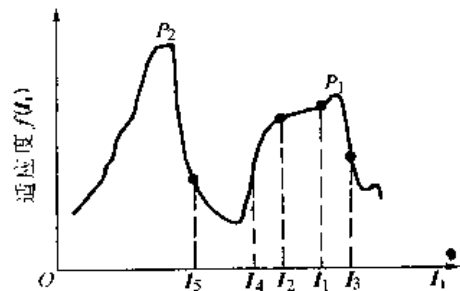


图 38.3-6 局部最优解的示例

4) 基因型的交叉 从被产生的  $N$  个个体中随机地只选择  $M$  组两个个体的配对,对各配对进行交叉 (crossover) 操作。进行交叉的概率叫做交叉率。

交叉是把两个个体的基因型以随机的方式,在其对应的基因位进行部分交换的操作。SGA 常用 1 点交叉 (one-point crossover) 来进行。图 38.3-7 是 1 点交叉的示例,基因型  $G_a, G_b$  作为一组被选中的配对个体  $I_a, I_b$ , 其基因型用  $G_a = \{11001010\}, G_b = \{11100111\}$  表示。

这时,将基因型在被随机选定的交叉位置切断。如果基因型为长度  $n$  的字符串,则可供选择的交叉位置有  $n+1$  个。在图 38.3-7 的示例中,交叉位置是由左第 4 和第 5 字符之间的位置。由交叉所生成的下一代个体  $I_{ab1}$ 、 $I_{ab2}$  的基因型  $G_{ab1}$ 、 $G_{ab2}$  用下式表示:

$$G_{ab1} = \{11000111\}$$

$$G_{ab2} = \{11101010\}$$

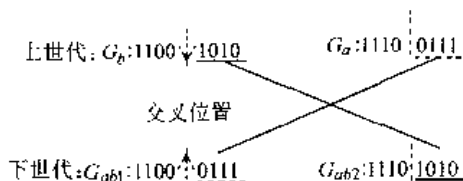


图 38.3-7 基因型的 1 点交叉示例

然后,个体  $I_a$ 、 $I_b$  由  $I_{ab1}$ 、 $I_{ab2}$  所取代。这就是基因型的交叉。基因型的交叉是世纪生物生殖过程的模拟。

由交叉所生成的个体  $I_{ab1}$ 、 $I_{ab2}$ , 是继承了上一代个体  $I_a$ 、 $I_b$  遗传信息的个体。如此处理,使生物集团中的基因型具有多样性,变得丰富多彩,实现基因型的进化。在初期阶段的生物集团中,本来具有多样性的基因型个体群,由交叉又会产生各种各样的新个体,整个生物集团的个体群并没有发生倾向性变化。另一方面,随着进化的进行,生物集团中的基因型逐渐出现某种倾向,即无论哪个个体,其基因型已无大的差别,此时由交叉产生的个体基因型与上一代个体具有相当大的相似性。也就是说,GA 的初期阶段是在搜索空间内进行全局性的调查;在掌握了其倾向性后,再进行更详细地搜索,求出其全局最大值。

5) 突然变异 把突然变异 (mutation) 发生的概率叫做突然变异率 (mutation rats)。突然变异有各种各样的方式,SGA 的突然变异是以突然变异率的概率方式,在每个基因位上进行变更操作,即以突然变异率确定发生突然变异的基因位后的操作,所进行的方法是若其基因位是 0 则变为 1,是 1 则变为 0。图 38.3-8 则是 SGA 的突然变异的示例。

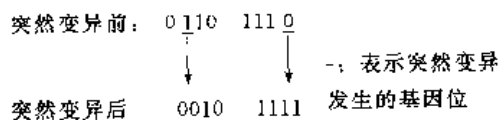


图 38.3-8 突然变异的示例

在图 38.3-8 中,即由突然变异率随机地确定发生突然变异,因此让其基因位上的字符发生相对变更。由突然变异的操作,会产生出仅由交叉而不可能产生的基因个体。从保持生物集团个体的多样性观点来解释,就是可能产生远离现有集团个体群的新个体(新搜索点),使搜索由局部最优解中脱出。应该注意的是,如

果突然变异率过大,将会使基因型失去由交叉所持有的上代遗传特征。一般,突然变异率在 0.1%~0.5% 为好。

6) 生物集团的评价 评价已生成的下世代生物集团是否满足进化仿真的评价基准,叫做生物集团的评价。评价基准一般由实际问题而定。GA 的典型结束评价基准如下:

- ① 生物集团中的最大适应度比某一设定值大。
  - ② 生物集团中的平均适应度比某一设定值大。
  - ③ 相对于世代进化次数,生物集团的适应度增加率仍在某值以下,在一定时间内没有大的变化。
  - ④ 相对于世代进化次数,达到设定的次数。
- ①和②是标准的评价基准。③表示个体进化一直处于低适应度环境下,生物集团处于搜索空间的局部最优点附近,搜索以失败而告终。④与③同样,表示其搜索也许已以失败而告终。

在满足评价基准的情况下,即可结束进化仿真。搜索成功时,仍把现生物集团中的个体表现型作为所求工程问题的解。在不满足评价基准的情况下,进化仿真反复进行。

### (3) 单纯型遗传算法的特征

SGA 具有以下三个基本操作:

- 1) 淘汰、增殖:以个体的适应度高低成比例地决定各个体在下世代生存的可能性。
- 2) 交叉:以随机性质的交叉率选择两个个体,并对其个体的基因部分进行交换。
- 3) 突然变异:以突然变异率随机地变更某基因位的值。

以上操作虽然很简单,但实用上对各类搜索空间的搜索却是有效的。SGA 中的各计算参数,如生物集团中的个体总数、交叉率、突然变异率等,至今仍无固定的设定模式,只能根据试算,或由经验给以设定。在此方面,有必要对 GA 算法作进一步的考察和研究。

GA 同传统的数学规划方法相比较,具有以下的特点和问题:

- ① 由若干个搜索点同时进行搜索,通过个体间的相互协调,具有可能避开其局部解的功能。
- ② 因为不需要使用评价值的微分,所以适用于不连续评价函数的求解问题。
- ③ 其具体的操作方法(淘汰、增殖、交叉和突然变异等)无一般的模式,需要根据实际求解问题,凭借经验和试算进行编程。
- ④ 大多数参数需人为地确定。

## 1.3 模式定理 (schemata theorem)

现阶段关于 GA 的理论解析还很不完善。由

J. H. Holland 提出的模式定理是 GA 的基本定理。这个定理为求出基因型中的基因排列, 在进化仿真中具有多大的生存概率, 提供了计算方法。本节简要地叙述模式定理。

简单地说, schemata (单数形式为 schema) 是夹杂着字符 \* 的字符列集合。例如, schemata  $S$  字符可表示为 4 个文字列,

$$\begin{aligned} S &= * 0011 * 01 \\ &= \{00011001, 10011001, \\ &\quad 00011101, 10011101\} \end{aligned} \quad (38.3-2)$$

这里, 字符 \* 有两种可能性, 可表示为 0 或 1。这时, schemata  $S$  的阶数 (order)  $O(S)$  以及构成长度 (defining length)  $\sigma(S)$  的定义如下:

$$O(S) = (\text{全长 } L) - (* \text{ 的个数}) \quad (38.3-3)$$

$$\sigma(S) = (S \text{ 中从最右到最左的 } * \text{ 字符间的距离}) \quad (38.3-4)$$

即式 (38.3-2) 中 schemata  $S$  的  $O(S) = 6$ ,  $\sigma(S) = 6$ 。

在无交叉和无突然变异、生物集团个体总数  $N$  一定的情况下, schemata  $S$  在世代交替中的平均增加率  $R$  由下式表示:

$$R = \frac{f(S)}{\frac{1}{N} \left\{ \sum_{i=1}^N f(G_i) \right\}} \quad (38.3-5)$$

式中  $G_i$  —— 个体  $I_i$  的基因型;

$f(S)$  —— 相对于  $S$  所表示的所有 schemata 的平均适应度。

式 (38.3-2) 的  $S$  用下式计算:

$$\begin{aligned} f(S) &= \frac{1}{4} \{f(00011001) + f(10011001) \\ &\quad + f(00011101) + f(10011101)\} \end{aligned} \quad (38.3-6)$$

式 (38.3-5) 中的右边的分母表示集团全部个体的平均适应度。设交叉率为  $p_c$  ( $0 \leq p_c \leq 1$ ), 那么因交叉 schemata 被切断的概率  $R_c$ , 由下式定义:

$$R_c = p_c \frac{\sigma(S)}{L-1} \quad (38.3-7)$$

在时刻  $t$ , 设定具有属于 schemata  $S$  基因型的个体数期待值为  $P(S, t)$ 。突然变异率为  $P_m$  时, schemata 不发生突然变异的概率是  $(1 - P_m) O(S)$ 。因为交叉和突然变异是相互独立的, 所以下式成立:

$$\begin{aligned} P(S, t+1) &\geq P(S, t) \frac{f(S)}{f} \left| 1 - p_c \frac{\sigma(S)}{L-1} \right| \\ &\quad \times (1 - P_m) O(S) \\ &\approx P(S, t) \frac{f(S)}{f} \left| 1 - p_c \frac{\sigma(S)}{L-1} - O(S) \right| \end{aligned} \quad (38.3-8)$$

式中  $\bar{f}$  = 平均适应度 =  $\frac{1}{N} \left\{ \sum_{i=1}^N f(G_i) \right\}$ ,  $N$  = 个体总数。

这里, 因突然变异率  $P_m$  比 1 充分小, 把  $(1 - P_m) O(S)$  用泰勒展开省略了二阶以上的项。不等号表示省略了因交叉和突然变异产生的由其他 schemata 混入的高阶成分。

式 (38.3-8) 叫做模式定理, 或者叫做遗传算法的基本定理。其定性表述如下: schemata  $S$  的构成长度  $\sigma(S)$  愈短, 阶数  $O(S)$  愈低, 平均适应度  $f$  就愈高, 在下世代生存的可能性就愈大。

以上的模式定理说明, 在基因型中存在适应度高、构成长度短的 schemata 的情况下, 其 schemata 被交叉切断的可能性低, 随着世代交替其数量将增加。这样的 schemata 被叫做积木 (building block)。在进化仿真中, 这样的积木有其若干种类的组合, 会产生具有高适应度的优秀个体。这就是所谓的积木假说。

模式定理虽然给出了调查某种 schemata 生存可能性的手段, 但是并不能解析在进化过程中新生成的 schemata 的动态。关于在进化仿真中变化着的基因型的数学解析和分析方法的研究, 将是今后的重要课题。

#### 1.4 遗传算法的有关操作规则和方法

GA 算法的流程同上一节所述的 SGA 大同小异。在处理实际问题时, 常常需要根据问题的特征, 考虑对其淘汰、增殖、交叉和突然变异等规则以及适应度的设定和定标等予以修正。以下介绍一些规则供参考。

##### (1) 淘汰、增殖规则的扩充

在 SGA 中, 选择下世代个体时, 由与各个体的适应度成比例的概率决定其个体被选中的可能性。假如完全服从这一基本方式, 那么在生物集团个体数少的情况下, 可能会产生现世代中具有最大适应度的个体, 作为下世代的个体偶然未被选中的情况; 生物集团个体的分布偏向适应度低的个体的情况。相反, 由于只选择具有高适应度的个体, 生物集团将陷入局部解的境地。为了回避这样的问题, 研究者提出了各种各样的扩充规则。

1) 概率淘汰的方法 在进化仿真中, 按照与各个体适应度成比例选择世代个体的情况下, 具有高适应度个体占主导地位时, 低适应度的个体作为下世代个体被选中的可能性就非常小。因此, 从局部解脱出也变得困难。由此, 我们可以不是单纯地采用相对于适应度的大小成比例关系选择方法, 而是把适应度按大小顺序排序, 决定其作为下世代个体被选中的可能性的方法。即使这样, 有时也会出现适应度的大小难以直接反映进化仿真的情况。

2) 优秀个体保存战略 优秀个体保存战略即把现世代个体中适应度最大的个体强制性地在下世代中给以保存的方法,如图 38.3-9 所示。在这种方式下,生物集团中的最大适应度的值,随着世代交替地进行是单调增加的。但是,当生物集团取得局部解时,最大适应度的值不仅不会增加,而且随着世代交替地进行,现时刻最大适应度个体的影响将会扩大,更难以从局部解中脱出。

3) 比例淘汰+交叉增殖的方法 把生物集团中的个体群按适应度的大小顺序排列后,将一定比例的下位个体无条件地淘汰掉,再将上位适应度高的个体配对进行交叉,产生新的基因型,实现增殖。此方式(如图 38.3-10 所示)在进化仿真顺利时,可高速地收敛于最优解。

但是,与优秀个体保存战略相同,当生物集团陷入局部解时难以从局部解中脱出。

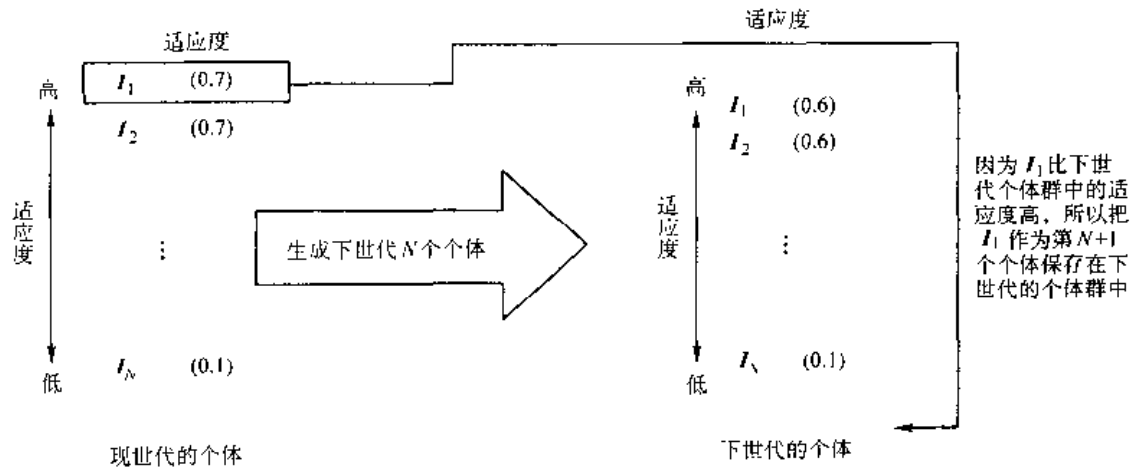


图 38.3-9 优秀个体保存战略

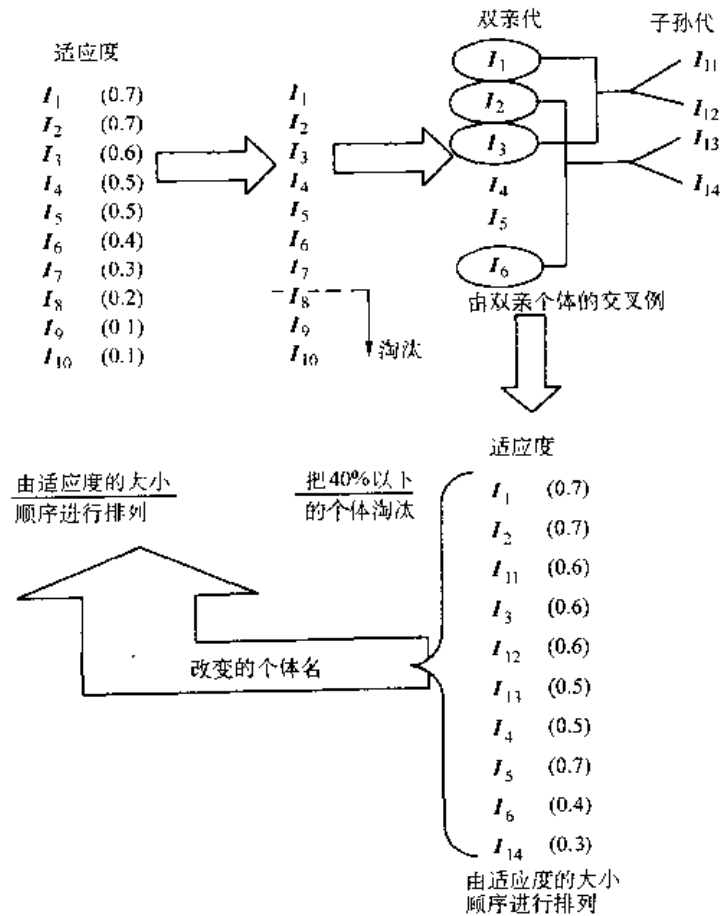


图 38.3-10 比例淘汰+交叉增殖的方式

总之,淘汰和增殖规则的原则是,尽可能地使高适应度个体(或者基因型的一部分)的 schemata 具有更高的生存可能性。但是,相对于具体问题,程序的最优淘汰、增殖规则又必须根据实际经验来确定。

### (2) 交叉操作的扩充

SGA 中使用的交叉是最基本的 1 点交叉方式。以下,介绍几种典型的交叉操作。

1) 2 点交叉 (two-point crossover) 图 38.3-11 所示的 2 点交叉方式不是把基因排列的基因型作为 1 列字符串,而是把最后的字符和首位字符连接成环状排列来处理。在环上随机地设定 2 点交叉位置,把环分割成两个弧,相互置换其相同的一部分,就构成了子孙基因的排列。

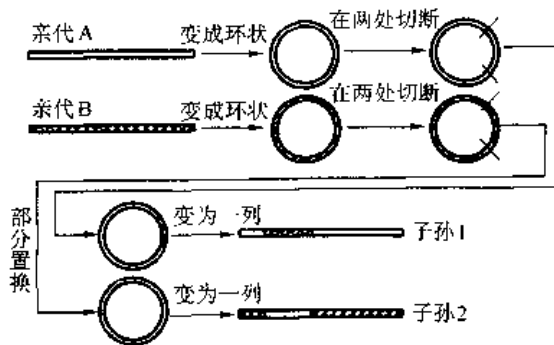


图 38.3-11 2 点交叉的示例

2) 多点交叉 (multi-point crossover) 多点交叉是 2 点交叉的扩充。多点交叉也与 2 点交叉相同,把基因的排列以环状来处理。然后,在若干个位置把环进行分割。这时,交叉位置的数目是奇数时,因为亲代 A 和亲代 B 的基因单元(弧段)相互等数置换有剩余,因此交叉位置的数目必须为偶数,即把多点交叉用 2 点交叉 ( $n$ -point crossover) 来表示时,  $n$  应为偶数。2 为奇数的  $n$  点交叉,在定义时需要丢弃  $n$  个交叉位置中的一个,或者需要把原来基因型的最后(尾部)再假设有第  $n+1$  个交叉位置来处理。

3) 分裂交叉 (segmented crossover) 分裂交叉是交叉位置总数可变的的多点交叉。其设定分裂置换概率为  $R_s$ 。例如,所谓分裂置换概率  $R_s=0.2$  表示对于长度为  $L$  字符串基因,其期望分裂长度为  $LR_s$ 。

4) 均匀交叉 (uniform crossover) 均匀交叉是在由亲代 A 和亲代 B 的基因型  $G_a$ 、 $G_b$  产生的子孙基因型  $G_{ab}$  时,  $G_{ab}$  的各基因占有亲代 A 的基因概率为  $p$ ,占有亲代 B 的基因概率为  $1-p$  的交叉。 $p=0.5$  产生的子孙基因型用下式表示:

$$G_a = \{11111111\} \quad G_b = \{10000000\} \quad (38.3-9)$$

$$G_{ab} = \{10011100\} \quad (38.3-10)$$

由式 (38.3-9) 和式 (38.3-10) 可以看出,子孙

基因型  $G_{ab}$  继承亲代 A 中  $L$  个基因中的  $p \times L$  个,亲代 B 的  $(1-p) \times L$  个基因。在对于问题难以判断用怎样的交叉合适的情况下,采用均匀交叉可以得到较好的搜索结果。

5) 混合交叉 (blended crossover) 混合交叉是在个体的基因型表示连续值的情况下,把两亲代的中间值作为于孙的基因型的交叉值的一种交叉。例如,在式 (38.3-11) 所示的亲代 A 的基因型  $G_a$  的值为 255、亲代 B 的基因型  $G_b$  的值为 1 的时候,把表示其平均值 128 的基因型式 (式 (38.3-12)) 作为亲代 A 和亲代 B 之间的子孙基因型  $G_{ab}$ 。

$$G_a = \{11111111\} \quad G_b = \{00000001\} \quad (38.3-11)$$

$$G_{ab} = \{10000000\} \quad (38.3-12)$$

除以上的几种交叉方式外,研究者提出了各种各样的交叉方式,在此不一一介绍。总之,使用的交叉方式是否合适,与要解决的问题、基因型的定义、适应度设定等密切相关。根据有关资料,1 点交叉和其他的交叉相比较其性能较差。关于各种交叉操作的有效性,以及交叉的数学定义和交叉对于搜索过程的影响等,还需要在计算理论和解析方面作深入的研究。

### (3) 突然变异规则的扩充

突然变异的作用是使由交叉产生的基因型具有多样性,即由突然变异产生仅由交叉不能产生的基因型,使搜索空间域大些。其次,在生物集团陷入局部解的情况下,具有脱出局部解的可能性。突然变异规则扩充的典型方式是,让突然变异率相对于进化过程的变化而变化。通常,生物集团的进化在顺利进行的状况中,突然变异率一定,抑制在低概率水平。生物集团适应度的增加率减少,能够判断生物集团陷入局部解的情况时,突然变异率比一般情况时要大,使其增加发生同现有集团基因型相异的搜索点,提高脱出局部解的可能性。

### (4) 引入适应度的定标

引入适应度的定标,主要是为了改善进化仿真初期和收敛时的淘汰功能。在生物进化仿真的初期,生物集团还不明确其方向性,各个体处于随机分布的状态。因此,偶然出现具有比其他个体适应度大的个体时,其方向虽然可能并不是真正的收敛方向,但生物集团的分布可能会出现偏向此个体方向的倾向,使生物集团被偶然性支配处于非常不稳定的状态,结果使生物集团不能到达最优点的可能性增高。为此,我们期望在进化仿真的初期阶段,相对于适应度的大小的淘汰不要太敏感。另外,进化仿真进入收敛的最后阶段,因为各个体的基因型已具有相当高的相似性特性,同进化仿真的初期阶段相反,期望选择优秀个体进行更深入的局部搜索,提高搜索精度,即相对于适应度的大小进行

严密的淘汰,为了实现这样的操作,需要随着进化仿真进行现行状况,改变对其适应度值的解释方式,这种改变对其适应度值的解释方式的操作,就是对适应度的定标。

所谓适应度的定标,如式(38.3-13)所示,由既定的设定方法,计算出某个体  $I_i$  的适应度的值  $f(I_i)$ ,代入函数  $G(\cdot)$  中,把求出的值  $f'(I_i)$  作为个体  $I_i$  的淘汰以及增殖计算的依据。

$$f'(I_i) = G(f(I_i)) \quad (38.3-13)$$

作为函数  $G(\cdot)$ ,有使用式(38.3-14)所示的线性函数的情况,也有使用式(38.3-15)线性函数的情况。

$$f'(I_i) = a \cdot f(I_i) + b \quad (38.3-14)$$

$$f'(I_i) = [f(I_i)]^k \quad (38.3-15)$$

式中  $a$ 、 $b$ 、 $k$ ——常数。

这里把式(38.3-14)的线性定标用图 38.3-12 表示。

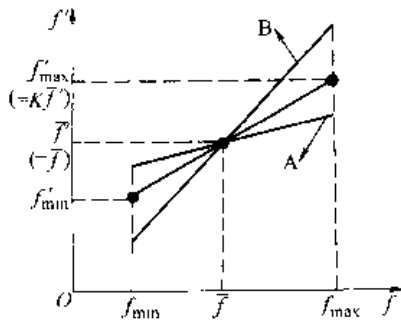


图 38.3-12 适应度线性定标

图 38.3-12 的线性定标中,  $f'$  的平均值  $\bar{f}'$  同  $f$  的平均值  $\bar{f}$  相等,即  $\bar{f}' = \bar{f}$ 。这样定标后,个体  $I_i$  的适应度的平均值  $\bar{f}'$  服从下式的基本淘汰规则时,其子孙在下世代生存的可能性是 1。

$$p(I_i) = \frac{f'(I_i)}{\bar{f}'} \quad (38.3-16)$$

式中  $p(I_i)$ ——个体  $I_i$  作为下世代个体被选中的概率;

$$\bar{f}' = \sum_{i=1}^N f'(I_i) / N, N \text{ 为个体总数。}$$

图 38.3-12 中的  $f'_{\max}$  决定了生物集团中具有最大适应度的个体在下世代中生存的个体数。若  $f'_{\max} = k \bar{f}'$ ,  $k=2$ , 即在下世代中生存的概率个体数为 2 个;  $k=3$ , 即在下世代中生存的概率个体数为 3 个。通常  $k$  的值在  $[1, 2]$  取值,  $k=1$  相当于不进行定标。

相对进化仿真进行的状况,改变式(38.3-14)中的系数  $a$ 、 $b$  的值,或者  $k$  的值,让直线的斜率发生变化,调解定标的效果。总之,进化仿真开始后的初期阶段,由图 38.3-12 中的直线 A 设定,适应度的值给与各个体的生存概率的影响将会降低。另一方面,生物集

团到达收敛时,以图中大斜率的直线 B 进行,适应度的值给与各个体的生存概率的影响将会增大。

至今,根据 GA 的适用问题的不同,有各种各样的扩充算法。这些算法是属于古典 GA,还是属于其他极不明确。现在,GA 的定义以及 GA 中操作方式的严密的定义尚不存在。在理论解析尚不充分的前提下,各种各样方法的尝试无疑是必要的。例如,由于同达尔文的进化论不同的进化论学说的出现,像不同观点进化论的讨论那样,有把 GA 以自由想象的方式进行扩充;也有开发非 GA 或者反 GA 的遗传算法的尝试。这些研究将促进以 GA 为代表的遗传方法不停留在经验的水平上,而在理论上逐渐趋于成熟。

### 1.5 多个体参与交叉的遗传算法

遗传算法是建立在模拟的基础上的,模拟把它们与自然界可能发生的过程联系在一起,模拟在具有解释功能的同时,也有一个不利的方面,那就是让问题变得模糊难懂,沿着模拟的道路走下去会阻碍那些不符合模拟所认可的方法的发展。遗传算法在应用时有其最根本的一个前提,即解的功能(适应值)与结构(编码)的相关性必须体现在:好的解之所以好是因为其具有好的基因(模式),而好的基因相互组合产生更好解的可能性很大。认清这样一个前提,我们回过头去考察以前遗传算法的运算过程,会发现交叉操作中只选两个个体参与是没有道理的,是受了模拟自然生物有性繁殖进化的限制。于是这里突破了模拟的限制,提出了多个体参与交叉的遗传算法,并将给出多个体参与交叉的遗传算法的模式定理,分析其对解群多样性的影响,初步解释其具有较高计算效率的原因,最后给出了一个算例验证了其具有较高的计算效率。

#### (1) 多个体参与交叉的遗传算法的简单描述

这里我们只描述与基本遗传算法相对应的多个体参与交叉的遗传算法。但请注意,几乎所有的遗传算法都可以转化为对应的多个体参与交叉的遗传算法形式。应用此遗传算法求解问题的准备工作与基本遗传算法大致相同,不过是控制参数中还应加上参与交叉的父代个体数目  $a$  和交叉后产生的子代个体数目  $b$  两个参数。这两个参数对遗传算法的性能有重要影响。

图 38.3-13 是与基本遗传算法相对应的多个体参与交叉的遗传算法框图。它与图 38.3-4 不同之处是用椭圆围起的那一部分,即交叉操作部分。从中我们也可看出,当提出了有多父代个体参与交叉的遗传算法概念之后,基本遗传算法便可理解为多个体参与交叉的遗传算法中参与交叉的个体数目  $a=2$  的特例。交叉方式可以有多种多样,这里我们提供一种方式。随机产生

$a-1$  个数值在 1 与串长  $L$  之间的整数, 这样参与交叉的  $a$  个父代个体都被分为  $a$  段, 相应子代个体的  $a$  段来自父代个体, 但不一定分别来自  $a$  个父代个体, 可能来自大与等于 2 个父代个体小于等于  $a$  个父代个体。这样可能产生新的个体数为  $a^a - a$  个, 可取子代个体数目  $1 \leq b \leq a^a - a$ 。例如选择了 3 个个体参与交叉, 随机选取两个交叉断点, 则可产生  $3^3 - 3 = 24$  个新个体, 子代个体数目可取在 1 到 24 之间。由此可见采用新的遗传算法可明显增加解的多样性, 从而提高计算效率。

(2) 多个体参与交叉的遗传算法的模式定理

**模式定理** 具有短的定义长度、低阶并且适应值在群体平均适应值以上的模式在遗传算法迭代过程中将按指数增长率被采样。

模式定理是遗传算法的基本定理, 它解释了遗传算法的有效性。若能证明多个体参与交叉的遗传算法的模式定理, 将会给这类遗传算法一个坚实的根基。

**定理 1** 模式定理对多个体参与交叉的遗传算法仍然成立, 而且通过调整参与交叉的个体数目对模式的短小性可有不同要求, 有利于快速求得 (近) 最优解。

**证** 假设在给定的时间步  $t$ , 一个特定的模式  $H$  有  $m$  个代表串包含在群体  $A(t)$  中, 记为  $m = m(H, t)$ , 用  $f(H)$  表示在时间步  $t$  含模式  $H$  串的平均适应值, 则交配池中含  $H$  的串数目为  $m(H, t) \cdot n \cdot f(H) / \sum_{i=1}^n f_i$ 。定义  $\delta(H)$  为模式  $H$  的长度, 即从第一个确定位到最后一个确定位的数目,  $l$  表示染色体串长, 则易知在一点交叉中, 模式  $H$  被破坏的可能性为  $P_d = \delta(H) / (l-1)$ , 那么生存概率为  $P_s = 1 - P_d$ ,  $a$  表示参与交叉的父代个体数目, 而多父代参与交叉的生存概率为  $P_s^{(a)} = (P_s^{(1)})^a$ , 因此在采用两个体一点交叉的遗传算法的下一代中模式  $H$  的数目为

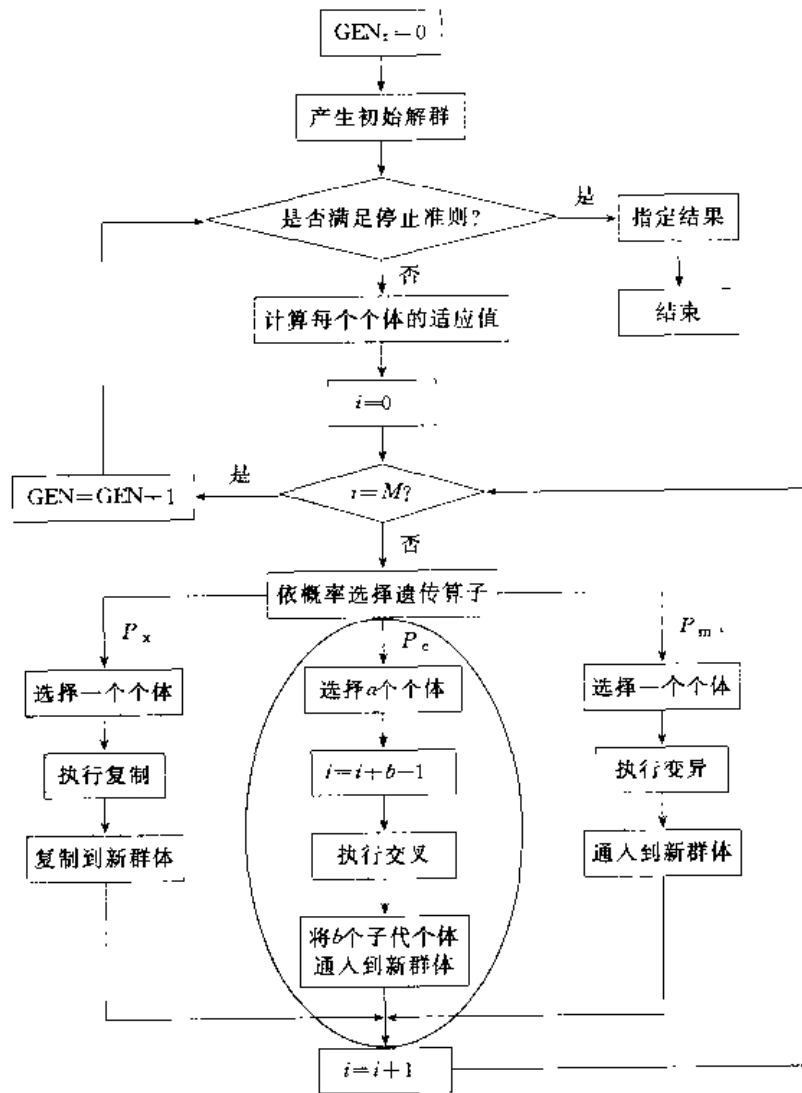


图 38.3-13 多个体参与交叉的遗传算法的框图

$$m(H, t-1) = m(H, t) \cdot n \cdot f(H) \times \left[1 - \frac{\delta(H)}{t-1}\right] / \sum_{i=1}^n f_i$$

而在多个体参与交叉的遗传算法的下一代中模式  $H$  的数目为

$$m(H, t+1) = m(H, t) \cdot n \cdot f(H) \times \left[1 - \frac{\delta(H)}{t-1}\right] / \sum_{i=1}^n f_i$$

由此可得定理1。

(3) 解群多样性的表示

我们可以从两个方面来理解解群的多样性，一种是解群的空间分布，显然从这个意义上来看，解群的空间分布越大，解群的多样性越强。因此我们用方差来描述解群的空间分布情况。

**定义1** 若第  $t$  代解群中的个体  $x_i^t$  由  $L$  个基因构成，即  $x_i^t = [x_i^{t(1)}, x_i^{t(2)}, \dots, x_i^{t(L)}]$ ,  $i \in \{1, 2, \dots, N\}$ ，定义第  $t$  代解群的平均个体

$$\bar{x} = [\bar{x}_1^{(1)}, \bar{x}_2^{(2)}, \dots, \bar{x}_L^{(L)}]$$

其中  $\bar{x}_l^{(l)} = \sum_{i=1}^N x_i^{t(l)} / N$ ，由此定义第  $t$  代的解群方差为

$$D_t = [D_t^{(1)}, D_t^{(2)}, \dots, D_t^{(L)}]$$

其中  $D_t^{(l)} = \sum_{i=1}^N (x_i^{t(l)} - \bar{x}_l^{(l)})^2 / N$ ,  $l \in \{1, 2, \dots, L\}$ 。

从定义1可以看出，方差  $D_t$  是  $L$  维的行向量，每一个分量表示出了解群在这维坐标上的空间分布。显然， $D_t^{(l)}$  越大，则解群在第  $l$  维坐标上的空间分布就越大。

方差仅反映出了解群的空间偏离程度，还不能完全刻画出解群的多样性。例如，解群  $\{1, 2, 3, 4, 5\}$  由五个个体组成，显然  $D=2$ ；而解群  $\{1, 1, 1, 1, 5\}$ ，方差  $D=2.56$ 。显然后者比前者的方差大，但前者比后者有更大的进化能力，为此我们引入描述解群多样性的第二个量——熵。

**定义2** 若第  $t$  代解群有  $Q$  个子集： $S_{1t}, S_{2t}, \dots, S_{Qt}$  各个子集所包含的个体数目记为  $|S_{1t}|, |S_{2t}|, \dots, |S_{Qt}|$ ，且对任意  $p, q \in \{1, 2, \dots, Q\}$ ,  $S_{pt} \cap S_{qt} = \phi$ ,  $\bigcup_{q=1}^Q S_{qt} = A_t$ ,  $A_t$  为第  $t$  代解群的集合，则定义第  $t$  代解群的熵如下：

$$E_t = - \sum_{j=1}^Q p_j \log(p_j)$$

其中  $p_j = \frac{|S_{jt}|}{N}$ ,  $N$  为解群中个体数目。

定义2告诉我们两点，第一点是当解群中所有个体都相同时，即  $Q=1$ ，这时熵取最小值  $E=0$ 。第二点

是当  $Q=N$  时，熵取最大值  $E=\log(N)$ 。解群中的个体类型越多，分配得越平均，熵就越大。对于十进制编码，熵的最大值为  $E_{max}^{10} = \log(N)$ ；对于二进制编码熵的最大值为  $E_{max}^2 = \log(\min(N, 2^L))$ 。

如果解群的方差和熵都很大，解群的情况如图 38.3-14a 所示，一般初始解群就是这个样子。当解群的方差很大，熵很小时，如图 38.3-14b 所示，解群集中在几个点上。解群的方差很小，熵很大时，解群集中在一个很小的区域中，如图 38.3-14c 所示。解群的方差和熵都很小时，解群收敛，如图 38.3-14d 所示，经过多代进化后，解群处于这种状态。

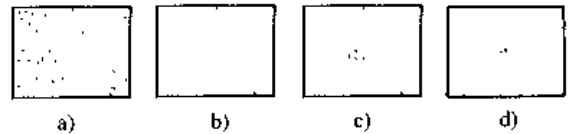


图 38.3-14 解群多样性与方差和熵的关系

(4) 多个体参与交叉的遗传算法对解群多样性的影响

多个体参与交叉的遗传算法的交叉操作可以描述为随机地从解群  $x_1^t, x_2^t, \dots, x_N^t$  中选出  $a$  个个体，依交换概率  $P_c$  进行基因重组运算，产生解群  $x_1^{t+1}, x_2^{t+1}, \dots, x_N^{t+1}$ 。Xiaofeng Qi 证明了当解群数目无穷，交叉操作保证每个坐标的边缘概率密度函数不变，即

$$g_{x_l^{(l)}}(x^{(l)}) = g_{x_l^{(l)}}(x^{(l)}) \quad (38.3-17)$$

其中  $l \in \{1, 2, \dots, L\}$ 。

**定理2** 无论交叉操作是两个个体参与交叉，还是多个个体参与交叉，都不改变解群的方差， $D_t = D_0$ 。 $D_0$  是初始解群方差。

证 由式 (38.3-17)，定理2显然成立。

**定理3** 若解群中每个个体有  $L$  个基因位，且初始解群中有  $Q$  个子集： $S_{1t}, S_{2t}, \dots, S_{Qt}$ ，对任意  $p, q \in \{1, 2, \dots, Q\}$ ，有  $S_{pt} \cap S_{qt} = \phi$ ,  $\bigcup_{q=1}^Q S_{qt} = A_t$ ,  $A_t$  为第  $t$  代解群的集合，若取两个个体参与交叉，则对任意  $m, n \in \{1, 2, \dots, Q\}$ ,  $x_i^t \in S_{mt}, x_j^t \in S_{nt}$ ，被分为两段满足

$$x_i^t \neq x_j^t \quad (38.3-18)$$

则经过交叉操作后会产生两个不同类型的个体。

证 由式 (38.3-18)，定理3显然成立。

**定理4** 若解群中每个个体有  $L$  个基因位，且初始解群中有  $Q$  个子集： $S_{1t}, S_{2t}, \dots, S_{Qt}$ ，对任意  $p, q \in \{1, 2, \dots, Q\}$ ，有  $S_{pt} \cap S_{qt} = \phi$ ,  $\bigcup_{q=1}^Q S_{qt} = A_t$ ,  $A_t$  为第  $t$  代解群的集合，若取  $a$  个个体参与交叉，则对任意  $C_1, C_2, \dots, C_a \in \{1, 2, \dots, Q\}$ ,  $x_i^t \in S_{C_1}, x_j^t \in S_{C_2}$ ，



...,  $x_i^a \in S_{C_a}$ , 被分为  $a$  段满足

$$\left. \begin{aligned} &x_i^1 \neq x_i^2 \neq \dots \neq x_i^a \\ &x_i^{1(1)} \neq x_i^{2(1)} \neq \dots \neq x_i^{a(1)} \\ &x_i^{1(2)} \neq x_i^{2(2)} \neq \dots \neq x_i^{a(2)} \\ &\dots \\ &x_i^{1(a)} \neq x_i^{2(a)} \neq \dots \neq x_i^{a(a)} \end{aligned} \right\} \quad (38.3-19)$$

则经过交叉操作后会产生  $a^a - a$  个不同类型的个体。

证 由于增加了式 (38.3-19) 的约束, 则交叉后可产生  $a^a$  个个体, 减去原有的  $a$  个个体, 得  $a^a - a$  个不同类型个体, 定理得证。

由定理 3、4 我们看出, 采用多个体参与交叉的遗传算法可明显增加解的多样性, 而且这种增加是指数增长的。例如取 3 个个体参与交叉, 则可产生  $3^3 - 3 = 24$  个新个体; 取 4 个个体参与交叉, 则可产生  $4^4 - 4 = 252$  个新个体。由定理 2 我们知道, 交叉操作不改变解群的方差, 交叉操作增加解的多样性事实上是增加了解群的熵。而采用多个个体参与交叉的遗传算法能呈指数快速性增加解群的熵。

定理 5 若解群中每个个体有  $L$  个基因位, 且初始解群中有  $Q$  个子集:  $S_{11}, S_{12}, \dots, S_{1Q}$ , 对任意  $p, q \in \{1, 2, \dots, Q\}$ , 有  $S_{1p} \cap S_{1q} = \emptyset, \bigcup_{q=1}^Q S_{1p} = A_1, A_1$  为第 1 代解群的集合, 若取两个个体参与交叉, 则对任意  $m, n \in \{1, 2, \dots, Q\}$  和  $i \in \{1, 2, \dots, Q\}$ ,  $x_i' \in S_m, x_i'' \in S_n$ , 满足

$$x_i' \neq x_i'' \quad (38.3-2)$$

$$x_i^{(1)} \neq x_i^{(2)} \quad (38.3-21)$$

则经过交叉操作,  $t \rightarrow \infty$  时, 解群中有  $Q'$  个不同类型的个体。

$$Q' = 2(L-1)C_Q^2 : Q \quad (38.3-22)$$

若子集  $S_k$  中个体的  $L$  位基因是由  $i_1, i_2, \dots, i_L \in \{1, 2, \dots, Q\}$  子集中的相应基因元素构成的, 则子集  $S_k$  所对应的  $p_k = \frac{|S_k|}{N}$  由下式可得

$$p_k = p_{i_1} p_{i_2} \dots p_{i_L} \quad (38.3-23)$$

当  $t \rightarrow \infty$  时, 解群的熵  $E$  满足

$$\lim_{t \rightarrow \infty} E = - \sum_{k=1}^{Q'} p_k \log(p_k) \quad (38.3-24)$$

证  $Q$  个子集中任两个子集中各取一个个体, 交换的位置有  $L-1$  个, 由定理 5 的条件式 (38.3-20) 式 (38.3-21) 可知, 一定产生两个新个体  $x', x''$ , 且  $x', x'' \notin S_q, q \in \{1, 2, \dots, Q\}$ , 所以交叉可产生  $2(L-1)C_Q^2$  个新型个体, 再加上原有的  $Q$  个类型, 得出式 (38.3-22)。由定理 2 得知, 交叉操作不改变每维坐标的基因元素, 显然式 (38.3-23) 成立。由熵的定义容

易得出式 (38.3-24)。得证。

定理 6 当  $t \rightarrow \infty$  时, 多个体参与交叉的遗传算法的熵与两个体参与交叉的熵相同。

证 任意数目个体参与交叉后产生的不同类型个体都可以通过两个体多次交叉后产生。易得定理 6。

从定理 5、6 我们还可以看出, 交叉操作不能保证解群的熵在  $t \rightarrow \infty$  时达到最大, 熵的变化与初始解群的分布有关, 当初始解群的所有个体都相同时, 交换操作不改变解群的熵。

用二进制编码时, 式 (38.3-19)、式 (38.3-21) 不易满足, 在基因数目  $L$  相同的情况下, 二进制编码产生的新个体类型数目比十进制少, 但并不意味交叉操作对二进制编码解群熵的提高能力小。因为对于十进制数, 往往采用多位二进制表示, 所以对同一问题, 二进制编码的基因位总是高于十进制的基因位。

(5) 一个算例

考虑下面测试函数的极小值:

$$f(x) = \sum_{i=1}^{10} x_i^2, \quad -5.12 \leq x_i \leq 5.12$$

虽然此函数是简单的平方和函数, 只有一个极小值, 但由于变量个数较多, 解空间庞大, 因此应用基本遗传算法求解的效率还是不高的。我们将用此函数比较基本遗传算法与多个体参与交叉的遗传算法的性能。为公平起见一些基本特征及控制参数设置如下:

二进制编码

轮盘赌选择

初始解群相同

解群规模  $N=50$

交叉概率  $P_c=0.8$

变异概率  $P_m=0.001$

$a=2, 5, 10$ , 相应  $b=2, 5, 10$

之所以如此设置第 7 条, 是因为随机选取多个个体参与交叉增加了计算量, 相应增加子代个体数目会平衡这种计算量的增加, 使得任一遗传算法计算一代的时间大致相同。图 38.3-15 所示为进化过程中最优适应度的变化, 图 38.3-16 为进化过程中平均适应度的变化。从中我们可以看到, 多个体参与交叉的遗传算法的确比基本遗传算法性能有较大提高。

(6) 讨论

本节提出了多个体参与交叉的遗传算法, 它的思路来源于对遗传算法本质的分析, 尽管没有任何模拟的基础, 我们通过一个算例验证了其具有较高的计算概率。在下文中我们将应用数学工具对多个体参与交叉的遗传算法进行初步分析。

多个体参与交叉的遗传算法的思路来源于对遗传算法本质的分析, 没有任何模拟的基础, 通过一个算例

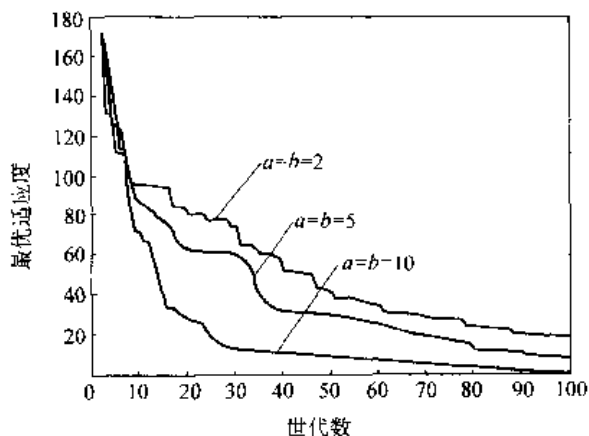


图 38.3-15 每代最优适应值比较

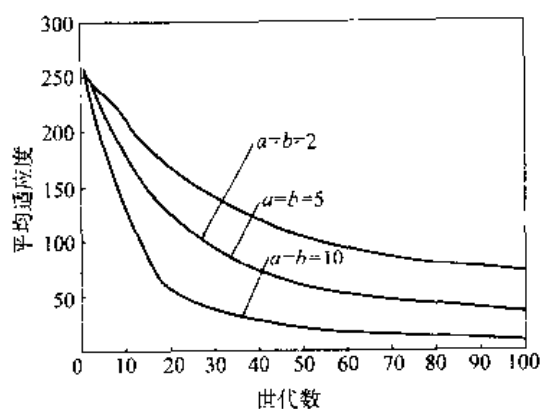


图 38.3-16 每代平均适应值比较

验证其是有效的。本节给出了多个体参与交叉的遗传算法的模式定理,并分析了其对解群多样性的影响,初步解释了高效性的原因。然而这里只是初步工作,还有许多问题有待解决。比如,①对全局收敛性的影响怎样?②参与交叉的父代个体数目取多少为好?③这里选取参与交叉的父代个体数目都是整数个,有没有可能选取实数个,例如选1.5个个体参与交叉,事实上许多应用就可看作少于两个个体参与交叉,最终寻得(近)最优解。④既然多父代个体参与交叉的遗传算法可快速增加解的多样性,那么解群数量是否可以减小?

### 1.6 多目标进化算法简介

如果决策是为了达到一个目标而进行各种策略的优化,那么,这类问题便称为“单目标优化问题”。如果所要决策的问题,不是为了实现一个目标,而是在为实现若干个目标的策略的优化、选择,那么,这类决策问题便称为“多目标优化问题”。

#### 1.6.1 传统多目标算法及其存在问题

##### (1) 多目标优化问题的相关概念

多目标优化 (MOP: Multiobjective Optimization Problem) 问题就是寻找满足由约束条件和目标函数组成的矢量函数的一组设计变量  $x_1, x_2, \dots, x_n$ , 使得决策者能接受所有的目标值。目标函数形成了对所设计的系统性能指标的描述。它的数学模型可以表达为

$$\begin{aligned} \min \quad & F(x) = (f_1(x), f_2(x), \dots, f_m(x))^T \quad (x = \{x_1, x_2, \dots, x_n\}^T) \\ \text{s.t.} \quad & g_i(x) \leq 0 \quad (i = 1, 2, \dots, p) \\ & h_j(x) = 0 \quad (j = 1, 2, \dots, q) \\ & x \in D \end{aligned} \tag{38.3-25}$$

通常称式(38.3-25)为一般多目标优化问题模型,其中  $n$  个变量  $x_1, x_2, \dots, x_n$  叫决策变量,由决策变量构成的矢量  $x = (x_1, \dots, x_n)^T$  称为决策矢量;  $m$  ( $\geq 2$ ) 个数值目标函数  $f_i(x) = f_i(x_1, \dots, x_n)$  ( $i = 1, 2, \dots, m$ ) 叫目标函数,由目标函数构成的矢量  $F(x) = (f_1(x), \dots, f_m(x))^T$  叫矢量目标函数;  $g_i(x) = g_i(x_1, \dots, x_n)$  叫约束函数,称

$$D = \left\{ x \in R^n \mid \begin{cases} g_i(x) \leq 0, i = 1, \dots, p \\ h_j(x) = 0, j = 1, \dots, q \end{cases} \right\} \tag{38.3-26}$$

为可行域。

**Pareto 最优解 (pareto optimal solution):** 对于问题式(38.3-25), 设  $x^* \in D$ , 如果不存在  $x \in D$  使得  $v = F(x) = (f_1(x), f_2(x), \dots, f_m(x))$  支配  $u = F(x^*) = (f_1(x^*), f_2(x^*), \dots, f_m(x^*))$ , 即  $\forall i \in \{1, 2, \dots, m\}, \neg \exists x \in D: F(x) \pi F(x^*)$

则称  $x^*$  为 Pareto 最优解。这个概念是经济学家 Pareto V. 于 1896 年引入的。

Pareto 解也称为非劣解 (non-inferior solutions)、容许解 (admissible solutions)、非支配解 (non-dominated solution) 或有效解 (efficient solutions)。有效解的定义表明,若  $x^*$  是有效解,则在可行域中找不出一个点  $x$  使  $x$  的所有目标函数值都比  $x^*$  的小,即至少存在一个  $x$  的目标函数值之一比  $x^*$  的相应目标函数值大。有效解的概念是多目标优化理论研究中的一个最基本的概念,求解多目标优化问题的目的就是求有效解。

**Pareto 最优解集 (pareto optimal set):** 对于给定的多目标问题  $F(x)$ , 所有有效解的集合称为有效解集,记为  $P^*$  (又称 Pareto 最优解集), 定义为

$$P^* = \{ x^* \in D \mid \neg \exists x \in D: F(x) \pi F(x^*) \}$$

**Pareto 前端 (pareto front):** 对于给定的多目标优

化问题  $F(x)$  和 Pareto 最优解集  $P^*$ , Pareto 前端 ( $PF^*$ ) 定义为

$$PF^* = \{u = F(x) = (f_1(x), f_2(x), \dots, f_m(x)) | x \in P^*\}$$

即非劣解向量在目标空间中构成 Pareto 前端, 决策者往往要根据 Pareto 前端显示的各目标值来进行最终的决策。

由于多目标优化问题的目标函数(指标)不是单一的, 造成最优概念的复杂化, 除有效解的概念外, 还有绝对最优解和弱有效解的概念。如图 38.3-17 所示为 MOP 问题的相关概念图例(以两目标为例)。

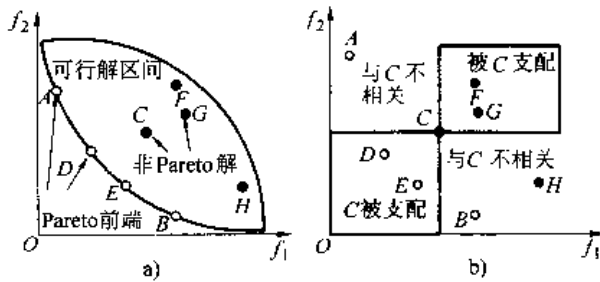


图 38.3-17 多目标优化相关概念图示  
a) 目标空间中 Pareto 最优解示例  
b) 目标空间中解之间的相互关系

### (2) 传统多目标优化方法

传统求解多目标优化问题的方法都是在搜索之前根据分析将多个目标组合成一个单目标, 然后求解。而这些组合参数的改变可能会导致不同的 Pareto 解。传统的具有代表性的多目标优化方法主要有: 加权和方法 (weighted sum method)、层次优化方法 (hierarchical optimization method)、字典排序法 (lexicographic ordering method)、权衡方法 (trade-off method)、全局准则方法 (global criteria method)、距离函数方法 (method of distance function)、目标规划法 (goal programming)、目标达到法 (goal attainment)、最小-最大法 (min-max optimum)、局部搜索法 (local search)、模拟退火法 (simulated annealing) 等。以上这些方法存在以下的缺点:

- 1) 一些方法, 如加权和方法等对 Pareto 前端的形状敏感, 要求搜索空间为凸。
- 2) 可能需要某些领域知识, 如权重的设定等, 而这些知识可能并不容易得到。
- 3) 容易陷入局部最优。
- 4) 每次运行只能得到一个解, 须多次运行才能得到 Pareto 近似最优解集。

近几年来, 进化计算成为研究多目标问题的有力手段, 通过进化计算可以处理复杂的搜索空间, 且每次运行都能得到多个解, 通过研究者的不断改进, 可以有

效地避免局部最优和保证解的多样性。

### 1.6.2 Pareto 多目标进化算法

近 30 年来, 人们对借助进化和遗传规律来解决问题的系统产生了越来越浓厚的兴趣。进化计算是模拟自然界的群体进化而逐步发展起来的。由于其所具有的本质并行性以及自组织、自适应和自学习等智能特征, 进化计算已成功应用到那些难以用传统的方法来进行求解的复杂问题中。进化计算包括了一类基于生物界的自然选择和自然遗传机制的计算方法, 如遗传算法 (genetic algorithm)、进化策略 (evolutionary strategies)、进化规划 (evolutionary programming) 以及 20 世纪 90 年代初期才发展起来的遗传程序设计 (genetic programming) 等。这里主要用到多目标遗传算法。

20 世纪 60 年代中期, 美国 Michigan 大学的 Holland 教授在 Fraser A S 和 Bremermann H J 等工作的基础上, 提出了适于变异、交配的位串编码技术, 后来将此算法用于自然和人工系统的自适应行为研究, 1975 年出版名著《Adaptation in Natural and Artificial Systems》, 并且将该算法推广应用到优化及机器学习等问题中, 这个算法就是遗传算法。

遗传算法是一种群体操作, 该操作以群体中的所有个体作为对象。该算法包含的基本算子是: 选择、交叉、变异。简单遗传算法有五个基本要素: 染色体表示 (参数编码); 初始群体的生成; 适应度函数的设计; 遗传操作设计; 终止准则。

遗传算法操作过程构成了简单遗传算法的基本结构。简单遗传算法的特点是: 采用轮盘赌选择方法; 随机配对; 采用简单交叉并生成两个子个体; 群体内允许有相同个体存在。

1896 年 Pareto 提出 Pareto 最优解的概念, 从而将多目标优化最优解从传统意义上的一个点扩展到一个集合, 也对多目标优化方法提出了新的要求。但传统的方法是从单目标优化方法上改进而成, 初始点只有一个 (或一群), 收敛到一个点上, 并不具有矢量优化的特点。

由于进化算法能够在一次运行中寻求多个 Pareto 最优解, 特别适用于求解多目标优化问题。20 世纪 80 年代中期 Schaffer 和 Fourman 对进化多目标优化算法进行了开创性的研究, 随后一些学者对其应用进行了研究, 近几年一些学者对多目标进化算法收敛到 Pareto 前端、小生镜和精华策略等方面进行了研究。另外一些学者则致力于开发新的进化方法。

#### (1) 多目标搜索中的主要问题及其改进

适应度分配和选择: 在单目标优化中目标函数和

适应度函数往往是等同的,而多目标问题中,适应度的分配和选择过程中要包含多个目标函数信息。在多目标进化算法中,按照对多目标在适应度函数中的表现方式,可以分为三类:各个目标单独考虑、经典的求和方法 and 直接应用 Pareto 占优的概念。

1) 通过变换目标进行选择 通过不断地在各目标间变换进行选择过程,而不是把各目标合并成一个适应度函数。用不同的目标函数来决定是否将个体复制和加入交配池,如 Schaffer 提出按比例地根据各个目标填充交配池; Fourman 的选择机制中将个体按照特定的目标排序进行选择;后来 Kursawe 将每一目标赋以一定的概率来决定在下一步的选择过程中该目标是否作为排序准则,其中的概率可以是自定义或随时间随机选择的。这种选择方法容易陷入极值点而得不到全局 Pareto 最优解,同时对非凸 Pareto 解前端很敏感。

2) 通过参数变换进行求和选择 采用传统的方法将各目标进行权衡,将多目标转移为带一定参数的单目标函数。参数在历次优化运行中是不变的,但在一次运行中可能是变化的。典型的方法是采用权重理论。因为每一个体按照特定的权重组合进行评价,使优化在多个方向上同时进行。这种方法同样对非凸 Pareto 前端敏感,从而限制了其应用。

3) 基于 Pareto 占优的选择 最早提出基于 Pareto 占优的个体适应度计算的是 Goldberg,他提出一种反复分级的过程:首先所有的非劣个体定为第一级并暂时从群体中移出,然后对余下的个体选取非劣点定为第二级,如此反复。个体的级别决定适应度值,此处的适应度值与全部个体有关,而其他方法中个体的适应度是单独计算而与其他个体无关的。许多学者都采用了这种方法。

尽管此方法理论上可以得到所有 Pareto 最优解,搜索空间的维数会对其效率产生影响。基于 Pareto 占优的选择方法是目前应用较为普遍的方法。

### (2) 精华策略

为了防止好的个体在取样或操作过程中的丢失,De Jong 提出将  $P_t$  中的最优个体保存到下一代群体  $P_{t+1}$  中。后来延伸到将  $b$  个最好的个体保留到下一代,这种策略就称为精华策略。在他的实验过程中发现:对于单峰函数精华策略可改善遗传算法的运作,而对于多峰函数却容易导致未成熟收敛。

不同于单目标优化问题,多目标进化算法中的精华模型更加复杂。在多目标问题中存在一个精华集而不是一个个体,精华集的规模相对于群体规模可能是非常大的。主要有两种基本的精华模型:

1) 一种是直接应用 De Jong 的思想,将  $P_t$  中的非

劣个体自动地保存到下一代群体  $P_{t+1}$  中。有时会对个体作一些限制,从非劣个体中选择  $k$  个到下一代。进化策略中的  $(\lambda + \mu)$  选择策略就属于这种精华策略。Rudolph 研究了基于  $(1+1)$  选择策略的多目标进化算法。

2) 另外一种精华模型是维持一个外部个体集合  $P$ ,其中的个体对应的决策矢量是当前所有解中的非劣矢量。每一代运行后,都会有一些个体加入或替换集合中的个体,这些个体可能是随机选择或根据其他准则,如个体在集合中存在的时间等。

### (3) 群体多样性

简单进化算法在进化中不断搜索具有好的适应度值的点,以较大的概率选择并繁殖它们。其结果是随着进化的进行,这些点的数目不断增多,最后整个群体收敛到其中一个点的附近,并将该点输出作为最优解。而在多目标优化中,如果要得到 Pareto 最优解集,就不能让群体收敛到一个小的区域,而应使群体保持整个设计空间中的均匀分布状态,这样才能保证得到的是一个解集而不仅仅是一个点。因而维持群体的多样性对确保多目标进化算法有效性是至关重要的。简单进化算法只收敛到一个点,并且常常由于选择压力(selection pressure)、选择噪声(selection noise)和操作中的破坏(operation disruption)而丢失较优的个体。为了解决这些问题主要有如下几种方法:

1) 适应度共享 (fitness sharing) Goldberg 和 Richardson 提出的适应度共享是目前最常用的一种小生镜(niches)技术,用以形成和维持一稳定的子群体。在特定的小生镜中的个体要共享可用的资源,在某一个体附近的个体越多,他的适应度值就要相应降低。其相临度用距离  $d(x_i, x_j)$  来定义,用小生镜半径(niche radius)  $\sigma_{share}$  表示。个体  $x_i \in P$  的共享适应度等于原有适应度除以小生镜计数(niche count):

$$F(x_i) = \frac{F'(x_i)}{\sum_{j \in P} s(d(x_i, x_j))} \quad (38.3-27)$$

个体的小生镜计数是该个体与群体内其他个体之间的共享函数(sharing function)值  $s$  的总和,通常把共享函数定义如下:

$$s(d(x_i, x_j)) = \begin{cases} 1 - \left( \frac{d(x_i, x_j)}{\sigma_{share}} \right)^a, & \text{若 } d(x_i, x_j) < \sigma_{share} \\ 0, & \text{其他} \end{cases} \quad (38.3-28)$$

距离函数  $d(x_i, x_j)$  可以定义为决策空间中的距离 ( $d(x_i, x_j) = \|x_i - x_j\|$ ) 或目标空间中的距离 ( $d(x_i, x_j) = \|F(x_i) - F(x_j)\|$ ),其中  $\|\cdot\|$  代表一定的距离评价。现有的大部分多目标进化算法

都采用适应度共享 (Hajela, 1994; Srinivas, 1994)。

2) 限制交叉 (restricted mating) 只有当两个个体在一定的距离范围内时(由参数  $\sigma_{\text{mnc}}$  定义)才允许交叉操作。与适应度共享相似,距离可以在决策空间或目标空间中定义。这种方法可以有效避免不适个体的形成,这种方法在多目标问题中并未广泛采用。

3) 隔离机制 (isolation by distance) 每一个体被赋予一个位置,根据距离对其进行隔离。主要有两种方法:一种是定义群体的空间结构,使空间小生镜在同一群体中进化;另外一种则是定义几个不同的群体,群体间只偶尔进行个体交换。如 Poloni 用多个小群体进行分布式进化。

4) 重新初始化 (reinitialization) 为了防止未成熟收敛,将整个群体或群体的一部分进行重新初始化, Golaberg 研究了使用小群体规模的遗传算法系统,每当遗传算法收敛,该系统就重新初始化, Goldberg 称其为序列选择 (serial selection)。Fonseca 定义了一种均匀多目标进化过程,其中在每一代的产生过程中都随机引入新的个体。群体的重新初始化在个体中引入多样性,提高了系统的性能。

5) 预选择机制 (preselection) Cavicchio 提出了预选择机制,规定只有在子代的适应度值超过其父代的情况下,子代才能替换父代而进入下一代。由于这种方式趋向于替换与其自身相似的个体,因而能够较好地维持群体的分布特性。

6) 拥挤机制 (crowding) 1975年, De Jong 提出了一种拥挤机制,其中新个体(子代)替换父代中相似的个体。在每一次进化中,从当前群体中选取  $1/CF$  ( $CF$  为拥挤因子)个个体组成拥挤成员,比较新个体与拥挤成员之间的相似性,决定是否用新个体取代拥挤成员中的相似个体。随着进化过程的进行,群体中的个体逐渐被分类。这种方法可以在一定程度上维持群体的分布特性。

通过聚类降低非劣解集中的个数,某些问题中, Pareto 解集可能会非常大。而从决策者的角度讲,获得过多的非劣解并无太大的意义,而且精华集中解的数目过多也影响了遗传的操作。一方面,由于  $\bar{P}$  要参与选择过程,过滤器中的解过多会降低选择压力,从而减弱了搜索的速度;另一方面,小生镜技术依赖于过滤器中的群体定义的网格分布均匀性,聚类过程中,将集合中的  $p$  个元素按照其相似性划分为  $q$  个群体,其中  $q < p$ 。有两种可以采用的聚类方法:直接聚类 (direct clustering) 和层次聚类 (hierarchical clustering)。直接聚类就是直接将  $p$  个个体分为  $q$  个集合。这里采用层次聚类法,按照平均关联理论 (average linkage method) 进行聚类过程。最初的 Pareto 解集中的所有

个体形成一基本的群体,随后的每一步中将两个集合合并成一个集合,直至达到所要求的集合数  $N$ 。合并集合的选取是按照相邻最近准则进行的,两个集合的距离定义为集合中两个个体间距离的平均值。分组过程完成后,从每一集合中选出代表个体,去除其他的个体。过程如下:

① 初始化群集  $C$ , 每一个个体  $i \in \bar{P}$  组成一单独的集合:  $C = \{U_i \in \bar{P} \mid \{i\}\}$ 。

② 如果  $|C| \leq N$ , 执行步骤①, 否则执行下一步; 计算所有集合两两间的距离, 两个集合  $c_1, c_2 \in C$  间的距离定义为集合中两个个体间距离的平均值

$$d_c = \frac{1}{|c_1| \cdot |c_2|} \sum_{i_1 \in c_1, i_2 \in c_2} d(i_1, i_2)$$

式中  $d$ ——两个个体  $i_1$  和  $i_2$  之间的距离。

③ 找出距离  $d_c$ , 最小的两个集合  $c_1$  和  $c_2$ , 合并组成一个集合:  $C = C \setminus \{c_1, c_2\} \cup \{c_1 \cup c_2\}$ , 回到步骤②。

从每一个集合中选取一个代表个体, 这里用重心法进行选取, 也就是说, 选取到本集合内其他个体平均距离最短的个体, 并除去其他个体, 剩下的所有个体组成后续循环的非劣解集。通过聚类, 既使 Pareto 解集中解的个数降至所要求的范围内, 同时也保持了原有解集的特性, 保持了解的多样性, 有利于避免局部最优。

④ 去除个体完成。

### 1.6.3 几种主要的多目标进化算法

下面选取几种主要的多目标进化算法进行比较, 对其主要特点和差异进行总结。

(1) Schaffer 的矢量评价遗传算法 (vector evaluated genetic algorithms: VEGA)

Schaffer 开发了名为 VEGA 程序的多目标优化程序, 其中包括了多判据函数。VEGA 系统的主要思想是将群体划分为相等规模的子群体, 每个子群体对于  $m$  个目标中的某个单个目标是“合理的”。对每个目标, 选择过程是独立执行的, 但交叉是跨越子群体边界的。进化计算过程中的适应度评价和选择过程在每一代的进化中都要执行  $m$  次。图 38.3-18a 所示为 VEGA 选择过程的图示, 其中对于某个目标较优的个体被选择复制, 加入到交配池中, 然后个体间进行交叉和变异。VEGA 归根结底仍是一种基于单目标的优先选择过程, 难以收敛到非劣解集。

(2) Hajela 和 Lin 的基于权重的遗传算法 (hajela and lin's weighting-based genetic algorithm: HLGA)

Hajela 和 Lin 提出了一种基于权重的遗传算法。它是以权重理论为基础, 并行的搜索多个解, 其权重并

不是固定不变的,而是在个体矢量中编码,这样在考虑不同权重组合的情况下对个体进行评价。图 38.3-18b 所示为 HLGA 的搜索过程。权重组合的改进是通过目标空间的适应度共享来实现的,所以进化计算同时优化解矢量和权重。同时 Hajela 和 Lin 还强调为了同时加快收敛速度和实现遗传搜索的稳定性,需要加入一定的交叉限制。

如前所述,权重方法只适用于解空间为凸的情况,这个问题也同样存在于 HLGA 中,但由于其简单性,仍被较多地应用。

(3) Fonseca 和 Fleming 的多目标遗传算法 (Fonseca and Fleming's multi-objective genetic algorithm; FFGA)

Fonseca 和 Fleming 提出了一种基于 Pareto 群体分级的选择方法,建立了个体的级别与当前群体中被该个体占优的染色体数目的关系。如果个体  $x_t$  在第  $t$  代,被当前群体中的  $p_t^{(1)}$  个个体占优,则其级别由下式确定:

$$\text{rank}(x_t, t) = 1 + p_t^{(1)} \quad (38.3-29)$$

图 38.3-18c 所示为假定的群体和相应的个体的级别,所有的非劣点的级别都定为 1。改进算法中将其他劣点的级别根据其所在位置的人群密度来考虑加适当的惩罚因子。这种对子块配置适应度的方法容易导致局部选择压力过大而过早不成熟收敛。为了避免这个问题, Fonseca 使用了一种基于共享机制小生镜技术来使群体均匀地分布在 Pareto 解集上。

FFGA 由于算法简单,易于实现,获得了广泛的应用。但它的效率依赖于共享因子  $\sigma_{share}$  的选择,且对共享因子非常敏感。

(4) Horn, Nafpliotis 和 Goldberg 的采用小生镜技术的 Pareto 遗传算法 (niched pareto genetic algorithm; NPGA)

Horn 等提出了一种基于 Pareto 占优的锦标赛选择方法,将锦标赛选择与 Pareto 占优的概念结合起来。与传统的仅仅在两个个体中比较不同的是,该方法中用到群体中一些其他个体来帮助判别占优性。该方法中的群体规模比其他的方法要大一些,使得在选择过程中出现的误差能够被小生镜克服。

图 38.3-18d 所示为 NPGA 的选择机制。图中白点代表的个体为竞赛的优胜者,因为它不被比较集占优,而另一个体被比较集占优。NPGA 的运行结果受共享因子和选择规模的影响。

Quagliarella 在此基础上提出了一种新方法,引入了选择过程中的占优准则,然后使用随机操作算子进行选择操作,但这样可能会导致选择出的个体仅在局部占优而不是在全局占优。由于这种方法不是对整个

群体进行选择,而只是在局部进行,所以该方法的速度很快,但使用中仍受到选择共享系数和选择规模的限制,需进一步研究。

(5) Srinivas 和 Deb 的非劣排序遗传算法 (nondominated sorting genetic algorithms; NSGA)

Srinivas 和 Deb 的支配排序遗传算法 (NSGA) 是基于个体的几层分级的。在选择执行前,群体根据是否支配其他个体被排序,所有非劣点被分成一类(带有一虚拟适应度值,它与群体规模成比例,用以为这些个体提供相等的复制潜力)。为维持群体的多样性,这些被分级的个体共享它们的虚拟适应度值。然后忽略这组分级的个体,考虑另一层非劣个体,继续上述过程直至群体中的所有个体被分级。

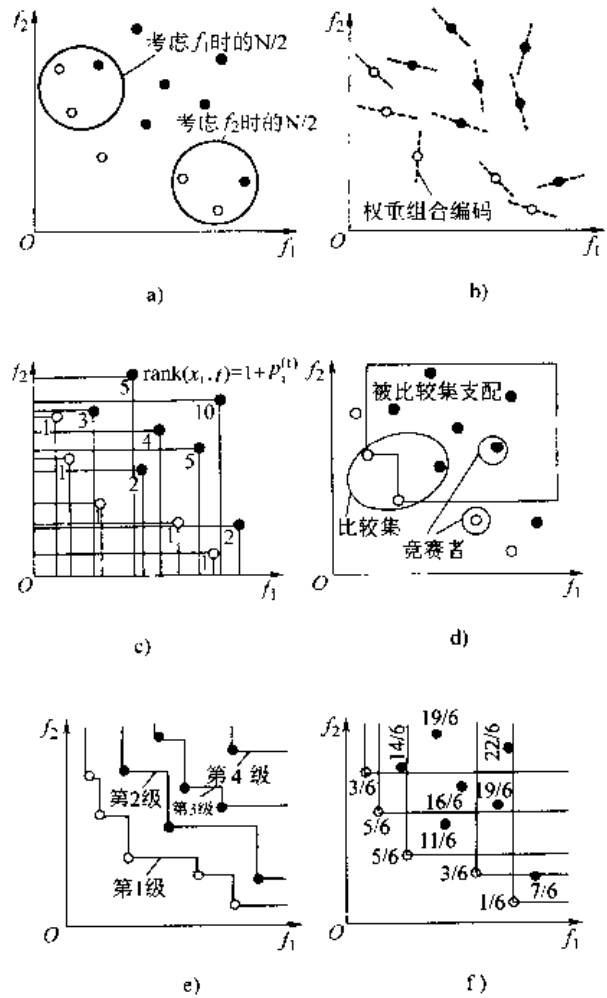


图 38.3-18 目标空间中六种不同的选择机制

- a) VEGA b) HLGA c) FFGA
- d) NPGA e) NSGA f) SPEA

图 38.3-18e 所示为 NSGA 的选择机制,其中低的适应度值对应于高的复制概率,与 FFGA 和 NPGA 不同的是,NSGA 的适应度共享在决策空间中进行。在 Srinivas 和 Deb 的研究中,适应度的设定与余下个体

的随机选择有关。

(6) Zitzler 和 Thiele 的增强 Pareto 进化算法 (strength pareto evolutionary algorithm: SPEA)

Zitzler 和 Thiele (1999) 提出了增强 Pareto 进化算法 SPEA, 它建立一外部群体  $P_e$  存储进化过程中所发现的非劣解, 使用了 Pareto 占优的概念来计算个体适应度值。SPEA 中群体成员的适应度仅由外部集合中的个体来确定, 而与群体中的个体是否占优于其他个体无关, 所有外部集合中的个体参与选择。为了保持群体的多样性, SPEA 中采用了一种新的基于 Pareto 的小生镜技术。

图 38.3-18f 为其适应度计算过程, 5 个非劣解矢量 (黑点表示) 把目标空间划分为几个矩形。这些矩形被看作小生镜, 目标就是使个体在这些网格中按一定的规则分布。与适应度共享不同的是, 小生镜不是以距离的形式, 而是以 Pareto 占优的形式来定义的。

### 1.6.4 扩展 Pareto 进化算法 (extended pareto evolutionary algorithm: EPEA)

本节提出扩展 Pareto 进化算法 EPEA, EPEA 将对前面所述各种技术进行综合应用, 并引进新技术, 以

提高多目标进化算法的效率和有效性, 得到 Pareto 最优解集。

#### (1) EPEA 算法过程

EPEA 的算法过程如图 38.3-19 所示。在每一次循环过程中, 外部的解集过滤器中的解集  $P$  都要更新, 如果超过解集过滤器的规模  $N$ , 要对其中的点进行非劣检查, 去除其中的劣点, 然后对  $P$  和  $P$  中的个体进行评价, 计算适应度值, 并从  $P+P$  中进行选择过程, 选择的个体进入交配池, 进行交叉和变异操作。下面分别介绍 Pareto 解集过滤器、小生镜和优秀解培育过程的操作。

#### (2) 用 Pareto 解集过滤器保存优秀个体

在遗传操作中, 可能在早期出现的某些染色体的适应度值要好于后来得到的最好染色体值。为了跟踪进化过程中的最好个体, 开辟一独立的位置用于保存这些曾经最好的个体, 称其为解集过滤器。在某些问题中, Pareto 解集可能会非常大。而从决策者的角度讲, 获得过多的非劣解并无太大的意义, 而且解集过滤器的大小也影响了 EPEA 的操作。一方面, 由于  $P$  要参与选择过程, 过滤器中的解过多会降低选择压力, 从而减慢了搜索的速度; 另一方面, 小生镜技术依赖于过滤器中的群体定义的网格分布均匀性。如果  $P$  中的个体

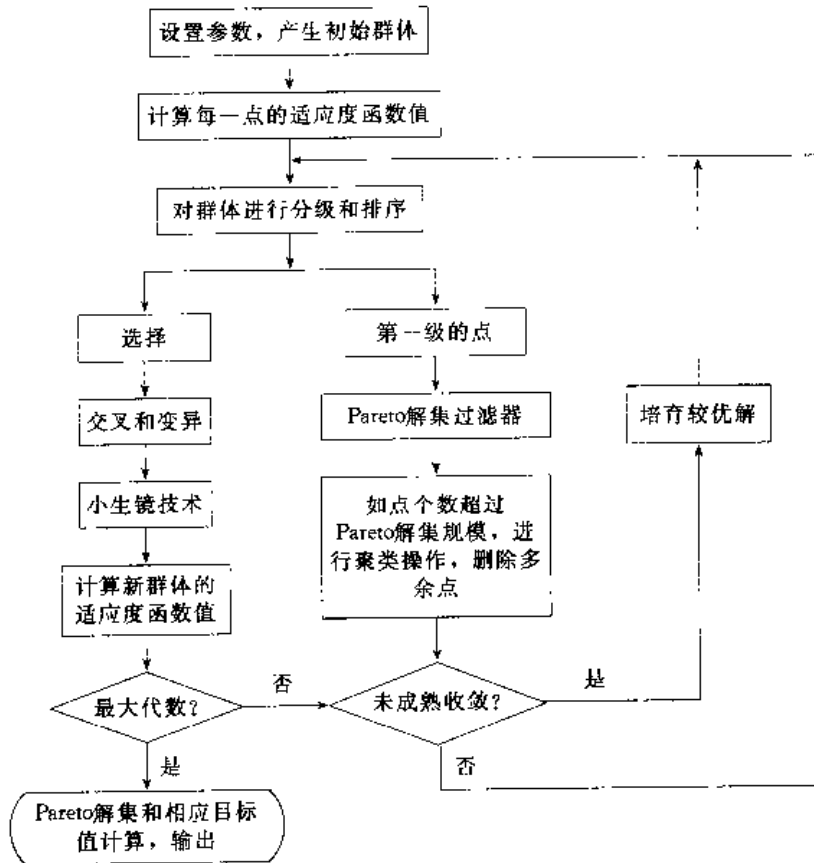


图 38.3-19 扩展 Pareto 遗传算法框架

分布不均匀, 适应度的计算会使搜索空间偏离到某一区域, 导致群体的不平衡, 因而如何在保持群体多样性的前提下精减过滤器中的解集是一个值得研究的问题。

基于以上的问题, 设计的解集过滤器需具有两个功能: ①对运行中产生的非劣解的存储和更新; ②运行过程中辅助对现有解和产生的新解的选择过程, 从而增加选择压力, 促使进化过程寻得更优的解。因而解集过滤器不仅仅是一个简单的存储器, 它需有一定的存储数目的限制, 这个数值反映了决策者所需的最终解的个数。如果经过遗传操作产生的新解不被其父代支配, 则将其与解集过滤器中的每一个解进行比较, 如果

新个体占优于过滤器中的个体, 则将其加入解集过滤器, 否则放弃。如果解集过滤器中的解超过了规定的规模, 则新解替换过程器中最拥挤处的点, 以利于群体的多样性。

为了表达解空间不同位置的拥挤程度, 引入  $d$ -维网格的概念。  $d$  是目标函数的数量, 对每次得到的新解, 在  $d$ -维网格中找到它的位置所在, 并记录各个网格中的解和解的数量。当过滤器已满时, 用新解替换最拥挤的网格中的解。如果新解既不支配过滤器中的解, 也不被过滤器中的解支配, 此时选择其中所处网格密度较小的解。图 38.3-20 为建立的 Pareto 解集过滤器的运作过程。

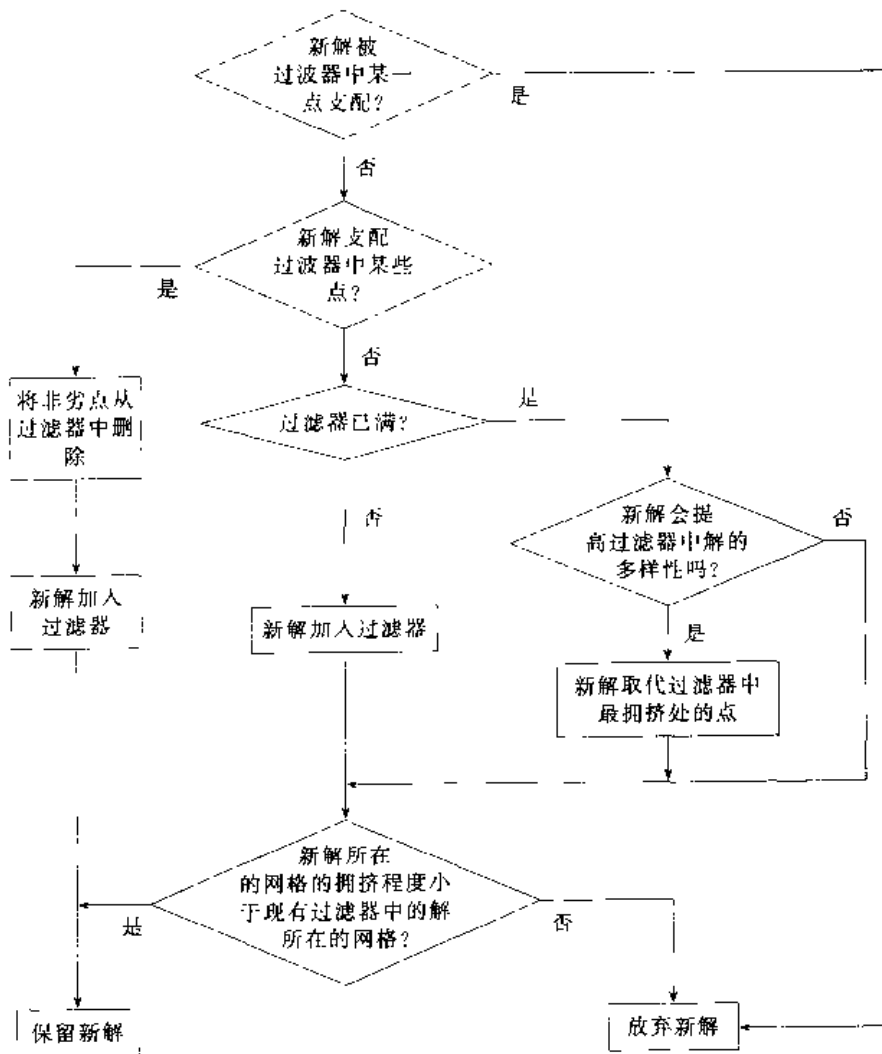


图 38.3-20 Pareto 解集过滤器运行过程

(3) 小生镜技术和优秀解的培育过程

在用遗传算法求解优化问题时, 我们既希望提高解的收敛速度, 又不希望看到局部收敛, 此处的 EPGA 中引入小生镜技术和优秀解的培育过程以权衡两者的

矛盾。

这里采用基于预选择机制的小生镜技术, 主要过程为: 父代中的两个个体进行交叉和变异操作后, 产生新的子代, 比较子代个体和父代个体的适应度值, 当子



代的适应度值超过父代时,用子代替换父代,否则父代进入下一代,这样,新的子代总是优于父代或至少与父代相似。采用基于预选选择机制的小生镜技术的主要作用是防止基因漂移,使群体中的个体沿 Pareto 解集保持均匀分布,同时也防止了群体中的个体向不利的方向进化。

在对产生的新解进行评价后,加入一个优秀解的培育过程,以加速过程的收敛。如果一个 Pareto 解在遗传操作过程中连续出现超过一定的次数(可以视为一种不成熟收敛的信号),则进行一次局部搜索过程,以提高解的多样性,改善的解进入下一代的遗传操作。局部搜索法从当前点出发,在当前点的一个邻域内搜索,如果邻域内的某个点比当前点的目标函数值更优,则用此点代替当前点继续搜索,直至搜索结束条件满足为止,培育过程为群体带入了新的信息,类似于变异操作的功能,与变异操作不同的是,优秀解的培育过程总是对 Pareto 解的改善过程。

### 1.6.5 算例

以下以几个算例来证明算法的有效性。

#### (1) 算例 1

求以下函数的最大值:

$$\text{目标函数: } \max f(x_1, x_2) = 21.5 + x_1 + \sin(4\pi \cdot x_1) + x_2 + \sin(20\pi \cdot x_2)$$

$$\text{约束条件: } -3.0 \leq x_1 \leq 12.1$$

$$4.1 \leq x_2 \leq 5.8$$

函数  $f$  如图 38-3-21 所示。

方法过程中选取如下的运行参数:

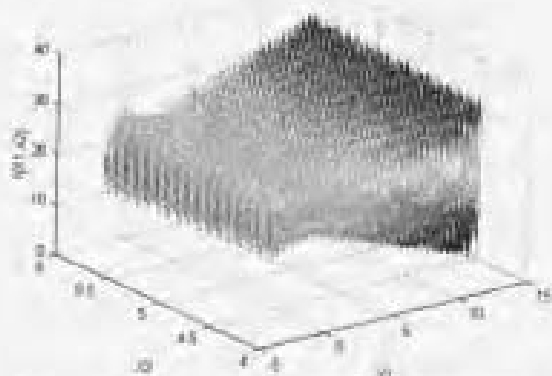


图 38-3-21 函数  $f(x_1, x_2) = 21.5 + x_1 + \sin(4\pi \cdot x_1) + x_2 + \sin(20\pi \cdot x_2)$  的图

群体规模: pop size = 20

交叉概率  $P_c = 0.4$ , 变异概率  $P_m = 0.005$

最大运行代数: MaxGenTerm = 200

得到的运行如图 38-3-22 所示,得到的最好染色体值为 38.850257 ( $x_1 = 11.6254$ ,  $x_2 = 5.7250$ )。用传

统的遗传算法运行 1000 代得到的最好染色体值为 35.477938。

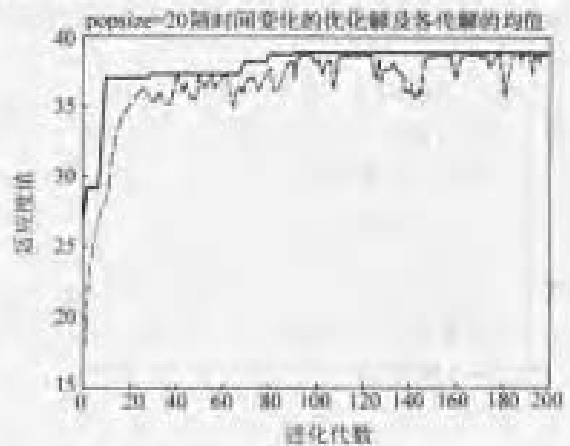


图 38-3-22 各代群体最优解及解的均值轨迹

#### (2) 算例 2

以 Schaffer 的经典算例进行检验和比较。

目标函数:  $\min f_1(x) = x^2$

$$f_2(x) = (x-2)^2$$

图 38-3-23 所示为两个目标函数的图形,图 38-3-24 中实线所示为多目标函数解空间的 Pareto 前端。

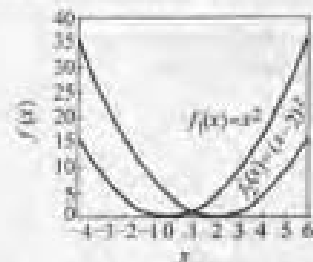


图 38-3-23 函数  $f_1(x)$  和  $f_2(x)$  的图形

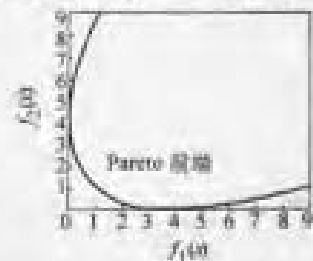


图 38-3-24 函数  $f_1(x)$  和  $f_2(x)$  的 Pareto 前端

进化参数:

群体规模: pop size = 30

交叉概率  $P_c = 0.04$ , 变异概率  $P_m = 0.005$

最大运行代数: MaxGenTerm = 200

解集过滤器数目: 100

仿真过程的设计空间为  $[-3, 3]$ , 如图 38-3-25 为进化过程产生的初始解, 经过 200 代的进化过程产

生图 38.3-26 所示的解集。

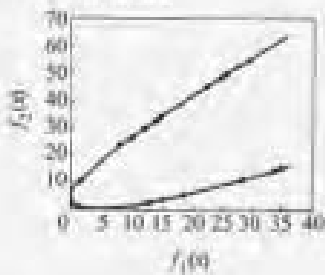


图 38.3-25 用 EPEA 产生的初始种群

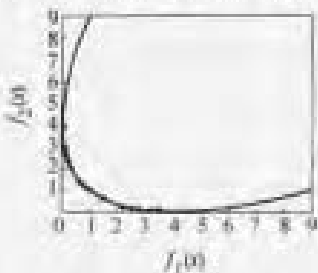


图 38.3-26 运行 200 代得到的种群

为了验证所加入的解集过滤器与优秀解培育过程的有效性,将其与 NPGA 算法进行比较,比较过程中采用相同的进化参数,分别对两种算法运行多次,比较运行过程中两者解的出现频率,其中对 NPGA 记录的是其最终解集,对 EPEA 记录的是最终解集过滤器中的解,比较结果如图 38.3-27、图 38.3-28 所示,可见进化的结果都能够在规定的进化代数内收敛到 Pareto 解集, EPEA 中落于真实 Pareto 解集区间的解的数量少于 NPGA,其解的分布也较 NPGA 分散,而 NPGA 的解向一侧偏离。

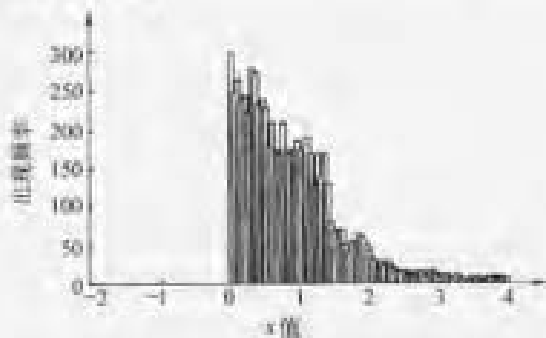


图 38.3-27 NPGA 运行过程中各解的出现频率分布

以上建立了扩展的 Pareto 进化算法过程,算例证明了算法的有效性,对于敏捷供应链重组决策中的多目标问题,也就是要求解满足决策者意愿的 Pareto 解,由于 Pareto 进化算法可以运行一次得到多个 Pareto 解,使决策者可以在多个目标间进行权衡,指导供应链的决策过程。

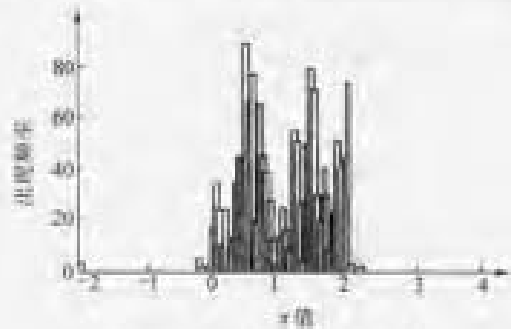


图 38.3-28 EPEA 运行过程中各解的出现频率分布

## 2 基于进化的健壮性设计方法

在复杂的系统初步设计过程中,影响质量的因素很多,如果对所有的参数进行全面试验,会耗费大量的时间及费用,这对于复杂的系统/过程来说是不实际的。

在选择试验设计方法时,需考虑到的一些要求为:在所研究的整个区域内能提供数据点的合理的分布;容许研究模型的适合性,包括对拟合不足的适合性;容许逐步建立较高阶的设计;提供内部的误差估计量;不需要大量的试验;不需要自变量太多的水平;确保模型参数计算的简单性。

这里,结合响应面法建立快速分析模块的特点,用序列试验设计方法结合改进的中心复合设计(CCD)作为设计特征样本点的方法。先用低阶的模型进行筛选试验,其结果用低阶的响应面模型来拟合,然后用方差分析(ANOVA)或回归分析来检验主要因素的显著性,将不显著的因素定位于中间水平,便于进一步调整,这样能减小问题的规模,确定感兴趣区域的范围(如图 38.3-29 所示的序列试验设计流程图)。

### 2.1 健壮性开发方法的基本思路

健壮设计的目的是为了在不消除产生变异的原因的情况下极小化这些变异造成的影响。早期的健壮设计主要考虑制造过程产生的变异的影响,近年来由于设计方法的发展,健壮性的概念已经进入设计阶段,即从设计阶段就考虑变异的影响,使设计出的产品或过程具有健壮性。

健壮性设计的基本原理是不仅达到性能目标,而且极小化性能变异,以此来提高产品或过程的质量。Taguchi 方法广泛应用于设计后期,实施参数设计和公差设计。这里介绍基于健壮性的产品及过程的开发方法将 Taguchi 方法应用到产品设计的前期阶段,将健壮性作为一个标准来帮助设计过程中的正确决策,同时尽可能减少问题的规模及设计后期的调整工作量。

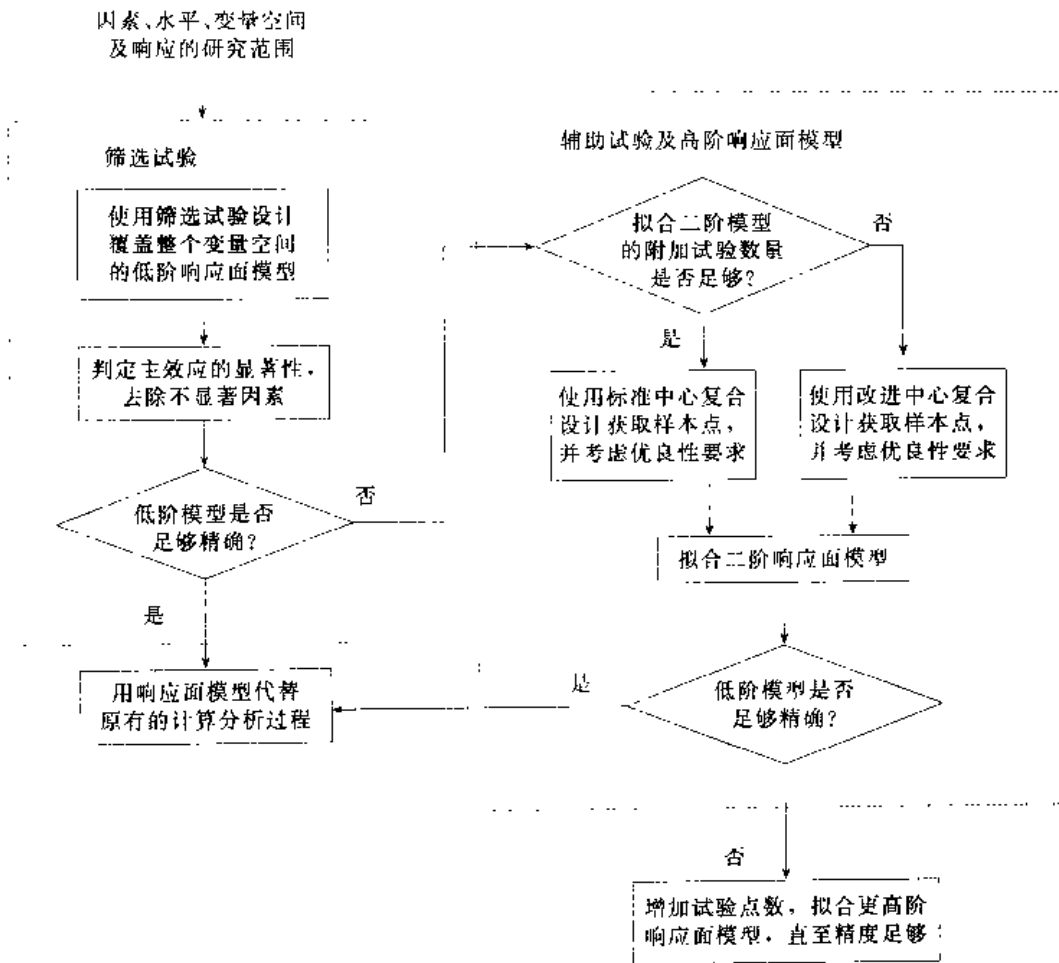


图 38.3-29 序列试验流程图

在健壮性产品及过程开发方法中，需要考虑三种因素：信号因素，即系统性能的目标；可控因素，即可以由设计者自由调整的参数；噪声因素，即不能被设计者控制的参数，如环境条件等。

变异的来源有两种，一是来自设计变量的变异，称为可控因素的变异；另一种是来自不可控因素，称为噪声因素的变异。

健壮设计问题按变异来源分为两类：

第 I 类：极小化由噪声因素（不可控因素）产生的变异；

第 II 类：极小化由可控因素（设计变量）产生的变异。

图 38.3-30 中描述了两种类型的健壮性设计问题。

图 38.3-30a 是第 I 类健壮设计问题的图示，图 b 是第 II 类健壮性问题的图示。其中可控因素  $x$  是能被设计者任意调整的参数；噪声因素  $z$  是不能由使用者调整的参数；信号因素  $M$  是系统输入，文中的推导过程假设系统输入是固定的。

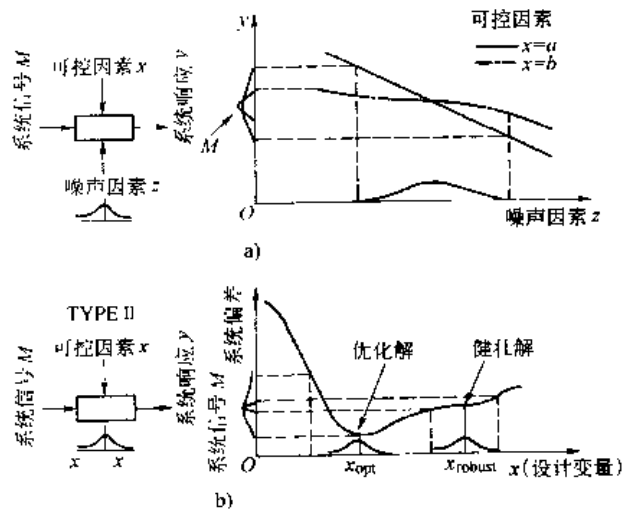


图 38.3-30 两种类型的健壮设计

a) 第 I 类健壮设计 b) 第 II 类健壮设计

在第 I 类问题中，系统的输出的变异的由不可控因素  $z$  引起。第 II 类问题与第 I 类问题的不同之处在于系统的输入中不包含噪声因素的作用，其性能的变异

仅仅由可控因素  $x$  在区域  $\pm\Delta x$  中的变异引起。

图 38.3-30 的右边部分表示了两种健壮性设计的不同概念,右上图为了第 I 类健壮设计;右下图为第 II 类健壮设计。经典的 Taguchi 的参数设计属于第 I 类的问题模式,对这种类型的问题的处理,是由设计者调节可控因素  $x$  来减小由噪声因素  $z$  造成的变异。图中的两条曲线表示当  $x$  处于两种水平  $x=a$  和  $x=b$  时性能变异作为噪声因素的函数的曲线。如果是一般的优化问题,要使性能尽可能接近目标  $M$ ,那么在  $x$  处于两种水平时都可行的,原因是它们的均值都是  $M$ 。但是如果引入健壮性概念后,当  $x=a$  时,系统性能在噪声因素  $z$  产生变异的情况下改变很大,而  $x=b$  时,变化较小。因此,  $x=b$  时系统从健壮性上优于  $x=a$ ,原因是  $x=b$  时减弱了噪声因素对系统的影响。

右下图中,为了便于说明,假设系统性能只是一个变量  $x$  的函数。如果仅考虑系统的性能优化,则应选取使系统性能取极值的最优点,图中  $x_{opt}$  表示优化后的最优点。但如果出于健壮性的考虑,需要寻找在可控因素产生变异时能使系统性能较稳定的点。显然,  $x_{opt}$  并不满足这个要求,满足要求的是  $x_{robust}$ ,当可控因素  $x$  在  $\pm\Delta x$  范围内波动时,在  $x_{robust}$  附近系统性能的变异很小,满足系统健壮性的要求。

虽然两种类型的健壮性从概念上来看有一定的区别,但从图 38.3-31 中可知,健壮设计的根本目标总是使系统响应的钟形分布(不一定是正态分布)的顶点(即均值)尽量接近质量目标,即优化性能均值,同时使钟形曲线的形状尽可能窄,即极小化方差。还要考虑可控因素或噪声因素的变异对工程约束条件的影响,这样得到的是一个带约束条件的多目标优化问题。这正是基于健壮性的产品或过程开发方法的出发点。

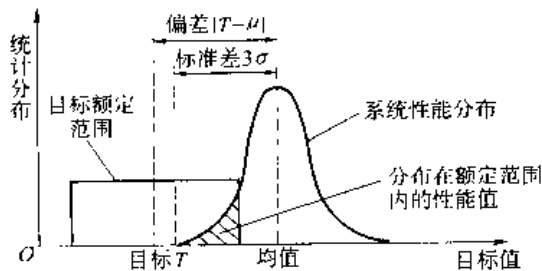


图 38.3-31 健壮设计中的性能统计分布

在前面几节讨论的各种技术的基础上,我们介绍了基于健壮性的产品及过程的开发方法。该方法包括下列技术:序列试验设计、响应面法、Pareto 多目标遗传算法。该方法的基本思路为:利用序列试验设计技术筛选出对系统响应影响较大的若干因素,设计出具有代表性的特征样本点,用响应面法建立系统的输入和

输出之间的近似映射函数,得到包含系统健壮性的目标函数和约束条件的多目标优化问题,最后利用 Pareto 进化算法求解多目标优化问题,得到使系统具有健壮性的全局最优解。

## 2.2 基于进化的健壮性设计方法的总体框架

基于健壮性的产品/过程开发方法的总体框架如图 38.3-32 所示。该算法包括四个主要步骤:

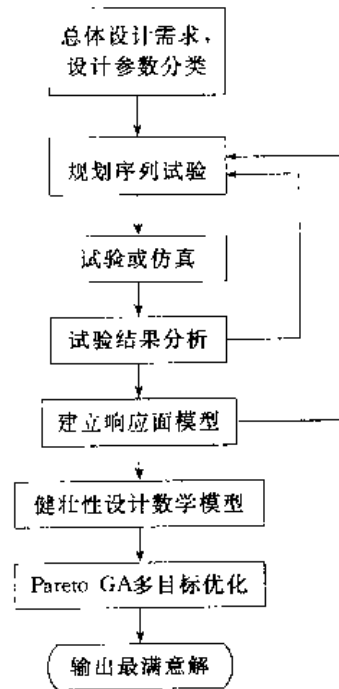


图 38.3-32 健壮性开发方法流程图

第 1 步: 规划模型参数;

第 2 步: 序列试验设计并建立响应面模型;

第 3 步: 根据健壮设计问题的类型建立目标函数和约束条件;

第 4 步: 用 Pareto GA 求解多目标优化问题,得到使系统具有健壮性的全局最优解。

以下将对这四个主要步骤的操作及数据流程作较详细的介绍。

第 1 步是对一个复杂的产品或过程进行分析,定义初始开发空间,规划健壮设计问题。将设计变量分组,可以由设计者直接调整的定义为可控因素,不能由设计者控制的归入噪声因素,同时限定这些因素的取值空间的范围。衡量系统性能的指标即系统响应也要在这一步中进行定义,并明确这一响应的目标是望大特性、望小特性还是望目特性,如果是望目特性,需指定目标值  $T$ 。为了减小问题的规模,需确定每一个响应的大致范围即感兴趣的区域。

由于复杂系统的设计过程中涉及到多种试验过程或仿真软件的计算,在第 1 步中需确定这些试验或仿

真的数据流及其输入参数及输出结果,并由此组成计算分析模块,该模块的作用是为下一步中设计响应面提供样本输出点。当响应面模型建立好后,就可由近似分析模块来取代该模块的位置,作为快速分析模块来实现原复杂耗时的计算工作。

第 2 步是一个序列进行的试验设计及结果分析过程。这个过程由规划序列试验、试验或仿真、试验结果分析、建立响应面模型几个模块组成。

先建立一低阶的筛选模型,设计少量的样本点,通过试验或计算机仿真,得到这些点处的系统输出作为响应值,对系统的因素(包括可控因素和不可控因素)进行分析,确定模型中的显著因素和不显著因素。用这些样本点和响应值拟合一个或一组一阶的响应面模型作为快速分析模块来取代原问题中的分析过程。如果用低阶模型的精度不够要求,需适当地增加样本点数来拟合二阶响应面模型,还可适当地增加阶数直到模型得到满意的精度。确定了模型的参数后,用回归分析或方差分析(ANOVA)来判别模型的显著性。

如果得到的模型是高阶模型,其中的交叉项还可以用来检验因素间的交互作用。交互作用的研究是很重要的,例如一个可控因素和另一个不可控因素之间的交互作用很强,在产品的使用中或过程的实际操作中不可控因素是不能由人为来调节的,就应该在设计中利用交互作用来调节可控因素,使不可控因素产生的影响最小。

整个序列试验设计和响应面拟合的过程是逐步提高拟合响应面精度的过程。在这一阶段,随着试验设计和响应面模型阶数的增加,在筛选试验中能尽可能地排除次要因素的干扰,逐步找出对系统响应的影响最显著的少量因素进行分析,问题的规模反面会越来越小。

第 3 步是由第 2 步的结果推导出健壮设计的数学模型。在该步中,主要是根据不同类型的健壮设计问题或它们的组合,建立健壮设计问题的目标函数及约束条件。

从响应面模型可以推导出表征质量特征的系统响应的均值和方差。针对前面提出的两种模型,这里介绍了不同的方法。对第 I 类健壮设计问题的模型,不可控因素(噪声因素)的变异是最重要的变异来源,此时使用一阶 Taylor 级数展开可得到响应的方差的近似表达。假设噪声因素独立,有

$$\mu_y = f(x, \mu_z) \quad (38.3-30)$$

$$\sigma_y^2 = \sum_{i=1}^k \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_{z_i}^2 \quad (38.3-31)$$

其中  $\mu$  表示均值,  $k$  表示响应面模型中噪声因素的个数,  $\sigma_{z_i}$  为每个噪声因素的标准偏差。

对第 II 类健壮设计问题的模型,可控因素的变异是最重要的响应变异来源,可得响应的均值和方差表达式

$$\mu_y = f(x) \quad (38.3-32)$$

$$\sigma_y^2 = \sum_{i=1}^l \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 \quad (38.3-33)$$

其中  $l$  表示响应面模型中可控因素的个数,  $\sigma_{x_i}$  为每个可控因素的标准偏差。当系统中同时存在可控因素和噪声因素的变异时,响应的方差由式(38.3-31)和式(38.3-33)组合而成。当然,用一阶 Taylor 线性展开会带来一定的误差,可改用二阶 Taylor 展开来提高精度,但这样会造成很大的计算量,在系统响应的变异不太大的情况下,用一阶 Taylor 展开已足够。

建立起响应的均值和方差表达式后,就可以建立健壮设计的目标函数了。健壮设计的目标有两个,转换为数学表达为

$$\min |T - \mu_y| \quad (38.3-34)$$

$$\min \sum_{i=1}^k \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_{z_i}^2 - \sum_{i=1}^l \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 \quad (38.3-35)$$

式(38.3-34)中  $T$  为响应的目标值,这种表示方法的目标函数适合于望目特性值。如果是望大特性值或望小特性值,需将式(38.3-34)作相应的修改。对于望大特性值的目标,式(38.3-34)变为

$$\max \mu_y \quad (38.3-34a)$$

对于望小特性值的目标,式(38.3-34)变为

$$\min \mu_y \quad (38.3-34b)$$

式(38.3-34)和式(38.3-35)即为形成的多目标优化问题的目标函数。由于可控因素和不可控因素的变异会对原问题的约束条件产生变异,需将约束条件进行转化。这里,以最坏情况分析法来研究,即假设系统的各参数同时处于最坏情况的组合。使用最坏情况的处理比使用概率设计方法更为保守一些,适用于要求较高的系统中。当设计变量的变异是以统计方式分布时,将原系统的约束条件

$$g_j(x, z) \leq 0 \quad (38.3-36)$$

转化为

$$E[g_j(x, z)] + n\sigma_{g_j} \leq 0 \quad (38.3-37)$$

其中  $n$  是由设计者根据所需要满足的可行性概率决定的常数。如果约束的变异呈正态分布,表 38.3-1 列出了不同  $n$  值对应的可行性概率,假设约束函数是正态分布。

约束条件的变异  $\sigma_{g_j}$  可由下式求出

$$\sigma_{g_j}^2 = \sum_{i=1}^k \left( \frac{\partial g_j}{\partial x_i} \right)^2 \sigma_{x_i}^2 + \sum_{i=1}^l \left( \frac{\partial g_j}{\partial z_i} \right)^2 \sigma_{z_i}^2 \quad (38.3-38)$$

表 38.3-1  $n$  与约束可行性概率的关系

$k$ 值	约束可行性概率 (%)
1	84.13
2	97.725
3	99.865
4	99.9968

如该变异不是统计量,式(38.3-37)转化为

$$E[g_j(x, z)] + \Delta g_j \leq 0 \quad (38.3-39)$$

其中  $\Delta g_j$  表示系统的变异,由下式近似得到

$$\Delta g_j = \sum_{i=1}^k \left| \frac{\partial g_j}{\partial x_i} \Delta x_i \right| + \sum_{i=1}^l \left| \frac{\partial g_j}{\partial z_i} \Delta z_i \right| \quad (38.3-40)$$

式中  $\Delta z_i$ —噪声因素的变异;

$\Delta x_i$ —可控因素的变异。

第4步是用 Pareto 进化算法来求解。

由前面提出的健壮性设计的分析中可以看到,以往的健壮设计均采用  $S/N$  比进行单目标优化,但一个目标值不足以表达统计量的特征,应同时用多个目标来实现。在实际应用中,往往会遇到工程约束的作用,这样最后的优化目标将是带约束条件的多目标优化设计问题,用传统的 Taguchi 方法是无法解决的。这里采用了能有效处理带约束多目标优化问题的 Pareto 遗传算法,利用矢量优化的概念,能得到全局最优点组合而成的 Pareto 最优解集,解集中的点包含了目标函数之间的各种权衡。算法包括 5 个基本算子:选择、变异、交叉、小生镜技术、Pareto 集合过滤器。该算法中设计了小生镜技术和 Pareto 集合过滤器,并建立了用于多目标优化的适应度函数,使用模糊罚函数法将带约束的多目标优化问题转换为无约束优化问题。为了求解具有混合变量的多目标优化问题,用离散变量圆整算子来处理混合变量,同时解决了如何在 Pareto 最优解集中加入决策者对各目标的重视程度来选择最满意点的问题。求解的结果是得到使系统具有健壮性的全局最优解。

### 2.3 基于进化的健壮性设计方法的说明

健壮性设计(Robust design)是质量工程中一个重要的概念和方法。从这个概念出发,发展出一种基于健壮性的产品及过程的开发方法,以产品质量或过程的健壮性为研究目标,采用序列试验设计技术筛选出对产品或过程影响较大的因素作为深入研究对象,用响应面法建立复杂产品或过程的输入参数及输出响应的快速分析模块,建立起以产品或系统的质量或健壮性为目标函数的数学模型,最后用多目标遗传算法优化得到产品或过程全局最优的健壮设计参数。该方法能得到产品或系统中可控因素或不可控因素产生变异时

使系统响应健壮的最佳参数组合,适用于以质量为主要目标的产品或过程设计和产品的概念设计及初步设计阶段。

采用试验设计技术(DOE)和响应面方法(RSM)有如下优点:

1) 试验设计技术可以通过试验或计算机仿真来研究不同设计参数的显著性。在设计复杂系统的初期,设计变量的数目往往相当大,此时有效地缩小研究的范围,将重点放在几个最关键的变量上是非常重要的。在试验设计的过程中可以实现这一目的。

2) 用响应面方法可以建立起复杂系统的性能空间(响应)和决策空间(设计变量)的直接映射关系,可以在一定精度上替代原复杂、耗时的分析模块,响应面模型可作为快速分析模块使用,这样更容易了解系统的输出和输入之间的关系。

3) 响应面法的优点在于简单、直观,并可用回归分析或方差分析研究检验模型的显著性及模型中系数的显著性,是一种较好的方法。从理论上来说,神经网络近似方法是一种很有前途的方法,只是因为神经网络的输出和输入之间是矩阵的关系,中间包含很多权值矩阵和阈值矢量,不容易写成简单的表达式,对下一步的推导健壮性产品及过程设计问题的目标函数不利,故没有采用。实际上,由神经网络的特性可知,神经网络除了可以建立系统输入和输出之间的映射关系,而且可以建立输入矢量和某一性能参数对设计变量的(偏)导数的映射关系。利用这一特性可以建立质量特性值的隐式表达的目标函数,具有较高的精度。同时,由于神经网络是一种高度非线性的映射关系,它更适用于非线性程度较高的模型。这一方法将作为新的发展方向提出,为后期的工作打下基础。

基于健壮性的产品及过程设计方法作为一个系统方法提出,是一种较通用的方法,适合于各种不同的学科及专业。本方法将设计过程中复杂耗时的模块用简单有效的近似模型替代,并能够找到使系统性能最健壮的全局最优点,其目的是在提高设计质量和计算效率的同时减少设计时间和设计费用。与传统的设计复杂工程系统的方法相比,有以下几个优点:

1) 健壮性设计方法将质量因素贯穿于产品设计的全过程中,使产品性能受外部环境及加工过程中扰动影响小,在各种条件下都保持较稳定的性能(质量)。

2) 基于健壮性的产品及过程开发方法使系统的响应能稳定在一个范围内而不受或少受设计变量的变异的影响,这样即使设计变量有较小的变化,系统的性能也基本不变。由此引申出健壮性设计方法的一个很大的优点,即特别适合于产品的设计过程。在成品的初

期设计阶段,有大量的因素要考虑,它们对系统的影响程度是未知数,需要有效的方法来研究哪些因素对系统响应的影响较大,哪些因素对系统几乎没有影响,将影响大的因素作为关键因素来进一步研究,确定这些因素的合适范围,使得在详细设计阶段,即使对这些参数作了较大的调整,也能保证成品的性能的变化很小,节省了重新设计的时间和人力,使设计具有柔性。

3) 使用基于健壮性的产品及过程开发方法设计出的产品,即使在制造或使用过程中受到内部或外部的干扰,使一些参数发生了改变,产品的整体性能仍能保持良好。在制造过程中,即使由于制造工艺或生产条件不合格导致产品的参数达不到精度要求,产品的性能仍能满足要求,这样就能用部分波动大、廉价的零部件、元器件来组装整机,只要参数搭配得合理,仍能使整机的性能十分稳定可靠,且产品成本低廉,最终使产品在市场上具有很强的竞争力。

4) 联合使用仿真和优化方法,可以广泛地研究整个设计空间,逐步将研究范围缩小到最感兴趣的区域,重点集中到对系统影响最显著的少数变量,最后得到满足健壮性的全局最优解。需要注意的是,得到的健壮解可能并不是最优的,但健壮解能使设计变量发生变异时系统响应的变异不大,同时仍能满足所有的约束条件,使产品或过程不仅实现了优化,同时能达到最优的质量特性,这是比传统的优化方法先进之处。

5) 与前面提出的利用灵敏度来实现健壮性的方法和 Taguchi 方法相比,本方法有较大的优越性。传统的基于灵敏度的健壮性设计方法是在每次迭代求得最优点后,再进行灵敏度分析,此时的导数信息只能用向外插值的方法预测变量在很小的范围内变化造成影响,只适用于设计后期即详细设计阶段,此时大部分参数已经确定的主要研究对象是局部和小的设计区域。Taguchi 方法也只能设计出对参数的较小变异不敏感的系统。 $S/N$  比方法,三次设计方法都适合于在设计后期设计产品的参数及容差。健壮性开发方法,由于引入了噪声因素作为设计变量,使设计空间进一步扩展。本文的方法适合于设计初期的概念设计及初步设计阶段,用来在较大的设计空间中寻找使系统具有健壮性的设计变量。

### 3 结构智能优化设计 —— 进化设计

结构优化设计理论已有近四十年发展历史,目前在一些重要的结构(如飞机结构)上已经得到一些应用。但是,造成结构优化设计的实际应用远远落后于理论研究这种现象还有一个很重要的原因,就是现有的

优化方法都是以传统的数学模型作为优化模型的。例如,准则法(包括力学准则法和理性准则法)、数学规划法(例如线性规划、非线性规划)以及两者的结合(即所谓的混合法)等静态优化方法都是基于代数方程模型的;最优控制理论中的庞特里亚金极大值原理、动态规划等动态优化方法是基于微分方程或差分方程模型的。由于这些传统数学模型描述能力和求解方法有相当的局限性,使现有的最优化理论和方法在实际应用中受到了很大的限制,存在许多有待解决的难题。这些困难包括:

#### (1) 局部最优解问题

复杂的优化问题可能存在多个解,其中,有若干个局部最优解(局部极大值或极小值),有一个全局最优解(全局极大值或极小值)。传统的解析寻优法只能寻找局部极值而非全局的极限。

#### (2) 维数灾难问题

虽然建立了各种可用的优化数学模型,但是,由于系统的复杂性,模型的维数高(设计变量数目和约束条件数目多),并且存在非线性等复杂因素,导致优化计算的工作量急剧上升,出现所谓的“组合爆炸”和“维数灾难”,造成求最优解的困难。

#### (3) 不确定性问题

在结构的设计中存在大量的不确定性因素(包括随机性、模糊性和未知性),传统的数学模型只能考虑确定性因素。

#### (4) 人的因素

结构设计属于软科学范畴,我们知道,软科学所研究和处理的一切问题必须有人的参与,必须充分利用人类专家的经验来解决问题。传统的数学模型是不能考虑人这个因素的。

作为人工智能科学和认知科学领域的最新发展,计算智能(Computational Intelligence,简称CI)当前正日益受到重视。一般认为计算智能包含神经网络(Neural Network,简称NN)、模糊系统(Fuzzy System,简称FS)和进化计算(Evolutionary Computation,简称EC)三个主要方面。由于这些新的理论和技术可以有效地解决人工智能研究中所遇到的局部最优解和组合爆炸等困难,因此,计算智能一出现就立刻受到世界各国科技界、企业界和政府决策机构的高度重视,计算智能被认为是对今后十年乃至21世纪的人类社会有重大影响的关键技术。

#### 3.1 结构智能设计的概念

揭示思维的本质,制造具有智能的机器,是人类一直所向往的目标。人工智能就是探索和模拟人的智能和思维过程的规律,并进而设计出类似人类某些智能

的自动机的科学。近年来,随着人工智能应用领域的不断开拓,传统的基于符号处理机制的人工智能方法在知识表示、处理模式信息及解决组合爆炸等方面所碰到的问题已变得越来越突出,人工智能学科正面临一个关键的发展阶段。

基于上述原因,寻求一种适合于大规模并行并且具有某些智能特征(如自组织、自适应、自学习等)的算法已成为人工智能学科新的重要发展方向,计算智能正是在这样的背景下产生的。计算智能的范围非常广泛,主要包括模糊系统、神经网络、进化计算、混沌计算和元胞自动机等,而以前三种学科为典型代表。这些学科有一个共同的特点:都是模拟某一自然现象或过程而发展起来的,并且具有适于高度并行以及自组织、自适应、自学习等智能特征,通过“仿生”(Imitating Life)或“拟物”(Simulating Physics)以使问题得到解决。

大自然是我们解决各种问题时获得灵感的源泉。几百年来,将生物界所提供的答案应用于实际问题求解已被证明是一个成功的方法,并且已形成一个专门的科学分支——仿生学(Bionics),因而,仿生成为计算智能一个特别引人注目的方向。例如,模糊系统与人工神经网络具有一个共同的目标:模仿人的大脑运行机制,从某种意义上讲,人工神经网络试图模仿人脑的“硬件”,而模糊系统旨在模仿人脑的“软件”。进化计算则是模拟生物进化和遗传现象的一种智能计算方法。

1994年在美国奥兰多召开的IEEE全球计算智能大会WCCI(IEEE World Congress on Computational Intelligence)是首次集人工神经网络、模糊系统、进化计算这三个最引人注目的仿生技术于一堂的计算机科学会议,会议明确提出了计算智能的概念,从此,计算智能作为一个独立的学科正式宣布诞生。

目前,计算智能正在向各个学科渗透。计算智能的出现和发展将为结构优化设计的研究注入新的生机和活力,两者的结合必将产生结构优化设计的一个新的分支,我们称其为“结构智能设计”,这就是说,

计算智能+结构优化设计=结构智能设计

### 3.2 结构进化智能优化设计

仿生的核心是进化(Evolution)。我们知道,自然界所提供的答案是经过漫长的自适应过程——进化过程而获得的结果。除了进化过程的最终结果,我们也可以利用这一过程本身去解决一些较为复杂的问题。这样,我们不必非常明确地描述问题的全部特征,只需要根据自然法则来产生新的更好解。进化计算(或演化计算)正是基于这种思想而发展起来的一种通用的问题求解方法。它采用简单的编码技术来表示各种复杂的

结构,并通过对一组编码表示进行简单的遗传操作和优胜劣汰的自然选择来指导学习和确定搜索方向。由于它采用种群(即一组表示)的方式组织搜索,这使得它可以同时搜索解空间内的多个区域。因此,用种群组织搜索的方式使得进化算法特别适合大规模并行计算。在赋予进化计算自组织、自适应、自学习等智能特征的同时,优胜劣汰的自然选择和简单的遗传操作,使进化计算具有不受其搜索空间限制性条件(如可微、连续、单峰等)的约束及不需要其他辅助性信息(如导数)的特点。这些崭新的特点使得进化计算不仅能获得较高的效率,而且具有简单、易于操作和通用的特性,而这些特性正是进化计算越来越受到人们重视的主要原因之一。

研究“结构进化智能优化设计”,就是把结构优化的数学模型转化为人工进化模型,通过进化算法来寻找最优解,从而达到优化的目的。因此,结构进化智能优化设计包括以下五个方面的研究内容:

- 1) 基于遗传算法的结构进化智能优化设计。
- 2) 基于遗传规划的结构进化智能优化设计。
- 3) 基于进化策略的结构进化智能优化设计。
- 4) 基于进化规划的结构进化智能优化设计。
- 5) 基于模拟退火算法的结构进化智能优化设计。

### 3.3 基于进化的桁架结构相位设计

GA算法由基因作为离散变量,对于组合最优化问题是有效的方法。在这里,列举结构设计组合问题的典型算例——确定桁架结构相位的问题,以说明GA的应用。

#### (1) 5节点桁架结构

以图38.3-33a的5节点平面桁架结构为例,对桁架结构的相位问题和GA的编码作以说明。相对于此基于结构的各杆件,定义其二进制编码为:杆件存在时用1,不存在时用0。因为此结构可能存在的杆件数为9,所以所有可能组合用9位二进制字符串编码。图38.3-33b表示了其对应编码的一例。

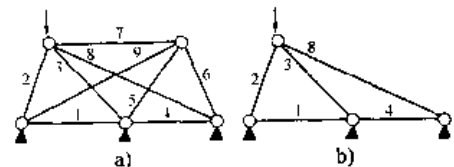


图 38.3-33 5节点桁架结构及其编码

a) 基本结构,染色体编码:[111111111]

b) 选择结构,染色体编码:[111100010]

相对于以上定义的桁架结构相位的问题的染色体进行适应度评价,即求解桁架结构相位在应力、位移约束条件下的最小重量问题。桁架结构的重量为 $f$ ,约束



条件为  $g_i \leq 0$ , 则其目标函数用下式定义:

$$F_j = \frac{1}{\phi_j} \text{ 或 } F_j = a\phi_j + b \quad (38.3-41)$$

式中  $j$ ——代表个体。

$$\phi_j = f_j + r \sum_{i=1}^N \max(g_{ij}, 0)$$

$$a = \frac{\phi_{avg}(c-1)}{\phi_{avg} - \phi_{min}}$$

$$b = \frac{\phi_{avg}(c\phi_{avg} - \phi_{min})}{(\phi_{avg} - \phi_{min})}$$

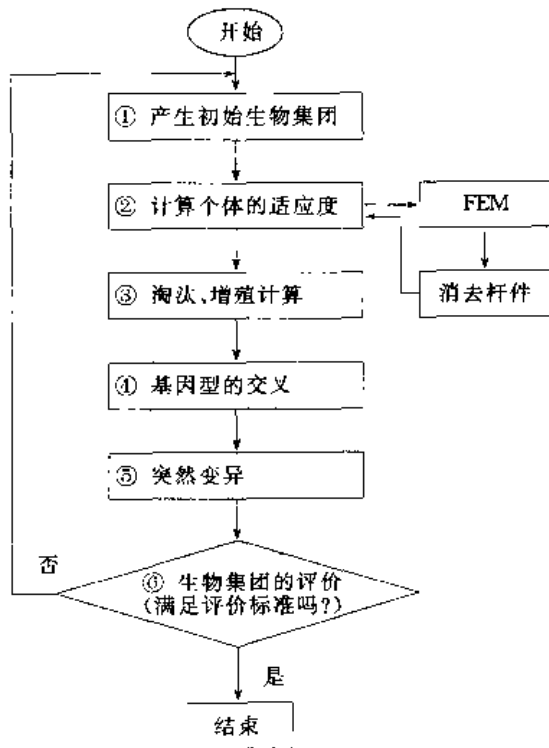


图 38.3-34 计算流程

其中  $\phi_{avg}$ 、 $\phi_{min}$  表示所有个体  $\phi_j$  的平均值和最小值;  $c$  是系数;  $a$ 、 $b$  为变换函数, 其目的是为了缓和初期生物集团中各个体适应值的急剧变化。

适应值需要通过结构解析获得。用 FEM(有限元法)对个体构成的结构进行解析。如果出现特异刚性矩阵的情况, 其适应值为 0; 如果产生具有某些杆件轴向力为 0 的结构, 则对其结构的个体编码进行修正。其计算流程如图 38.3-34 所示。

### (2) 9 节点桁架结构

使用以上的编码法和计算方法, 再来讨论图 38.3-35 所示的 9 节点桁架结构的相位求解问题。载荷节点(节点 8)有以下的位移约束条件

$$u_8 \leq u_0 = 0.015 \text{ mm} \quad (38.3-42)$$

杆件的弹性模量为  $2.06 \times 10^7 \text{ Pa}$  ( $21000 \text{ kgf/mm}^2$ ); 所有杆件的断面积相同。由此, 计算  $f$ 、 $\phi$ 、 $F$ 。GA

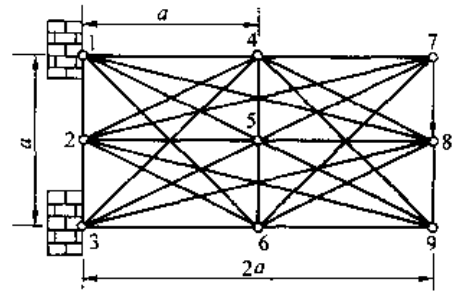


图 38.3-35 9 节点桁架和其基本结构

计算的约束条件为满足以下三个约束条件之一即可。

- 1) 集团高位个体的 20% 为同一染色体时。
- 2) 进行 5 世代进化仿真, 并未出现好解时。
- 3) 进化仿真超过 100 世代时。

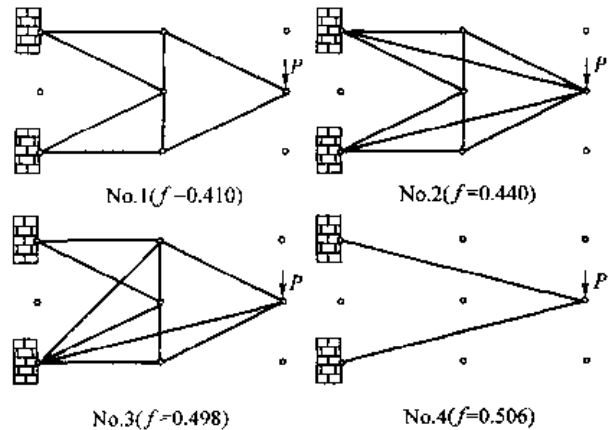


图 38.3-36 最优以及次优桁架解

图 38.3-36 是以个体数  $N=70$ , 淘汰比例  $P_t=50\%$ , 交叉率  $P_c=50\%$ , 突然变异率  $P_m=1\%$  计算所得的结果。最优解为 No.1, 可选次优解为 No.1~No.4。这里所讲的次优解, 是其评价值为最优解的 80% 以上的意思。如此, GA 解法的一大优点是, 即使得不到最优解, 也可得到许多次优解。

## 3.4 基于进化的结构非线性强制振动解法

### (1) 问题的提出

在各种机械约束条件下, 用数学规划法求解非线性问题的解已被广泛应用。但是, 问题的目标函数存在多峰性时, 其搜索得到的常常不是最大值(或最小值), 即存在陷入局部最优解的缺点。再之, 相对于非线性振动系统, 在同一激振外力和激振频率下, 常常存在着复数解(非线性方程的初始条件依存性问题)。因此, 在数学解法以及复数解之间的关系不明确的情况下, 要同时求出这复数解, 给最优计算模型的构筑带来较大的困难。另外, 即使作为工程问题的近似解来考虑, 因为数学规划法的搜索过程不仅过分依赖于搜索初始值, 而且需要目标函数和

约束函数的可微性条件,使得问题变得复杂或难以解决。相对于这类问题,因为遗传算法是以复数个初始值开始,仅用集团体搜索点的目标值评价进行搜索,所以有可能求得其近似最优解。这里以1自由度Duffing型非线性方程共振域复数解的求解为例,介绍GA适用于此类问题的解法。同时,探讨解的设定、问题的模型化、适应函数的定义和算法等问题,并用数值计算的结果检验解法的有效性。

(2) Duffing 非线性方程

作为解析模型,考虑在大振幅强制振动下的两端固定的梁。梁两端被固定,由梁的挠度在梁内产生张力;大振幅时,会出现3阶硬化型非线性刚度。这样的非线性强制振动系统难以得到其严密的数学解,只能使用数值积分的方法求出其数值解。把下式

$$m\ddot{x} + c\dot{x} + k(x + bx^3) = P\cos\omega t \quad (38.3-43)$$

中各系数设定为:  $m=2.56, c=0.32, k=1.0, b=0.05, P=2.5$

设定其主共振域解为

$$x = A\cos(\omega t - \psi) \quad (38.3-44)$$

把式(38.3-44)代入式(38.3-43),略去  $\cos\omega t$  项的影响,可得到图 38.3-37 的近似频响曲线。相对于共振域可  $\eta = \omega/\omega_n$  ( $\omega_n$  为线性固有振动圆频率),存在3个解。其中  $b$  是不稳定解,用通常的时间积分难以求得。只可能得到  $a$  和  $c$  两个稳定解。在此,共振域的频率比,  $\eta = 1.6$ , 在特定的初始条件

$$(x_{i_0}, \dot{x}_{i_0}) = (10.0, 0.0)$$

$$(x_{b_0}, \dot{x}_{b_0}) = (11.0, 0.0)$$

下,分别用RKG法积分,可得到图 38.3-38 的时间响应曲线  $a$  和  $c$ 。可以看出,其  $a$  和  $c$  稳定解由于初始条件的不同,其在图 38.3-37 频响曲线上的振幅和相位也不同。这样的具有3阶正刚性的Duffing系统的共振,其频响曲线向右弯曲。定性归纳其特性如下:

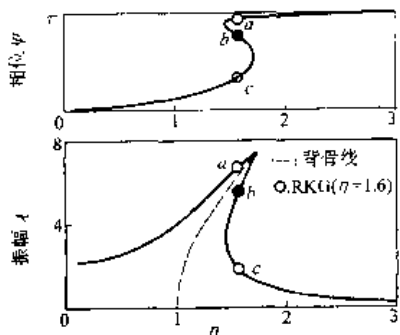


图 38.3-37 频响曲线

- 1) 相对于激振频率存在复数解的频率范围。
- 2) 大振幅振动  $a$  和小振幅振动  $c$  的周期与激振

周期相同。

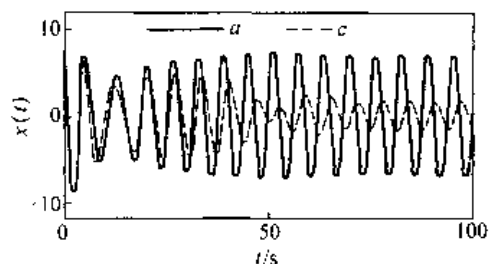


图 38.3-38 不同初始条件下的时间响应 ( $\eta=1.6$ )

3) 3个解中,位于  $a$  和  $c$  的中间的解  $b$  是不稳定解 稳定解  $a$  在上侧,  $c$  在下侧。

(3) 最优化问题的模型

这样的非线性问题的解析不存在一般性的方法。这里,利用最优化方法进行近似解析。根据以上的分析和图 38.3-37 所示 Duffing 系统的主共振特性,使用已知的信息,探讨同时求解复数解(稳定解)的最优化模型。设定下式

$$x_i = u_i \cos\omega t + v_i \sin\omega t + p_i \cos 3\omega t + q_i \sin 3\omega t \quad (i=1,2), (t > t_0) \quad (38.3-45)$$

为问题的解,其稳定主共振谐波解的振幅  $A_i$  和相位  $\Psi_i$  用下式表示:

$$A_i = \sqrt{u_i^2 + v_i^2} \quad (38.3-46)$$

$$\Psi_i = \arctan(v_i/u_i) \quad (38.3-47)$$

两个解中按振幅大小顺序,将其振幅和相位定义为  $A_0$  和  $\Psi_0, A_1$  和  $\Psi_1$ 。两个解必须在满足式(38.3-43)的微分方程前提下,构筑最优化模型。

设计变量:

$$X = \{u_1, v_1, p_1, q_1, u_2, v_2, p_2, q_2\}^T \quad (38.3-48)$$

目标函数:

$$J = m_1 J_1 + m_2 J_2 \quad (38.3-49)$$

式中

$$J_i = \sqrt{\frac{\sum (m\dot{x}_{i_j} + c\dot{x}_{i_j} + k(x_{i_j} + bx_{i_j}^3) - F_j)^2}{(m-k)}}$$

$$x_{i_j} = u_i \cos\omega t + v_i \sin\omega t + p_i \cos 3\omega t + q_i \sin 3\omega t$$

$$F_j = P\cos\omega t, t_j = j\Delta t$$

$$i = 1, 2; k = t_0/\Delta t, t_0 = 2N\pi/\omega_n$$

$t_0$  为目标函数开始计算的时刻。

约束条件:服从于前面所述的复数解的振幅和相位特征,列出以下条件:

1) 关于振幅的不等式关系

$$g_{11} = [\sqrt{4(\eta^2 - 1/3)/\beta}] - A_0 < 0 \quad (38.3-50)$$

$$g_{12} = A_c - A_0 < 0 \quad (38.3-51)$$

2) 关于相位的不等式关系

$$g_{21} = \Psi_0 - (\pi/2) - \Psi_1 < 0 \quad (38.3-52)$$

$$g_{22} = \Psi_0 - (\pi/2) < 0 \quad (38.3-53)$$

(4) GA 算法的适应度和流程

在这里,把 GA 用于前面所述的具有复杂解空间的复数求解问题。以上的最优化问题具有以下特点:把复数解作为最优化问题的设计变量;把不确切的信息作为不等式约束条件。

1) 适应度函数 同时求解复数个 Duffing 型非线性强制解的问题,在上一节中给出了其模型。为了把此最优化问题转换成适用于 GA 的离散化最优组合问题,最终需要转换成关于适应度函数的最优化问题,这里转换成无约束最优化问题,把外点惩罚函数作为适应度函数。

$$f(X) = J + g[\max(g_{1i}) + \max(g_{2i})] \quad (38.3-54)$$

2) 计算流程 使用适应度函数和图 38.3-39 所示的流程探索最优解。以下将使用的流程作一简单的说明。

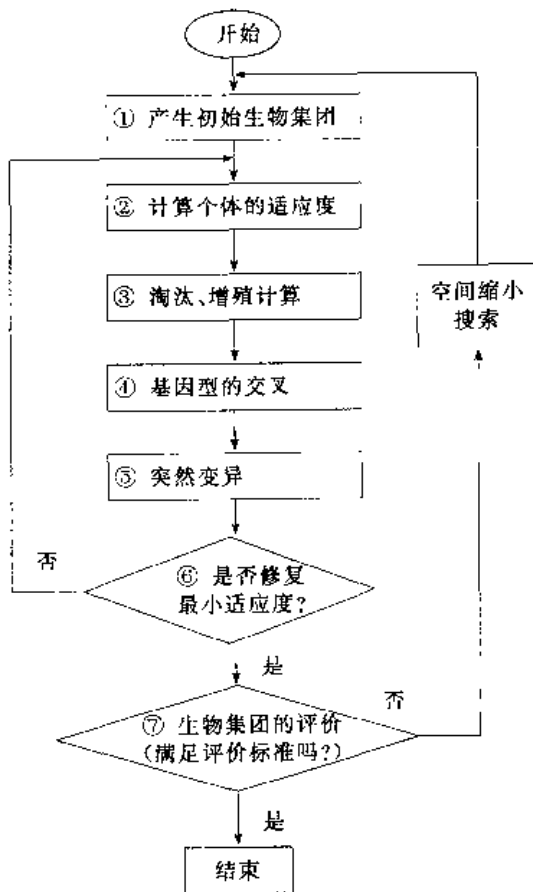


图 38.3-39 计算流程

① 二值化(Birth of Strings)。把连续最优化问题的设计变量用二进制数表示成离散变量。

② 增殖(Reproduction)。淘汰使用了轮盘赌选择+优良保存战略。一般地说,轮盘赌淘汰选择从计算

效率和内存容量等方面考虑,因个体集团不可能设定太大,所以某个体基因在集团中是被选择或被淘汰完全取决于淘汰压的概率波动。总之,作为最优化计算的流程,即使好不容易地出现了高评价的染色体,其遗传信息仍存在从集团中消失的可能性。因此,在增殖处理中,用同适应度成比例的概率对各个体的子孙进行淘汰处理后,把现在集团中适应度最高的个体作为下世代个体无条件地予以保存,此即优良保存战略。

③ 交叉(Crossover)。用 1 点交叉法进行处理。

④ 突然变异(Mutation)。对应于所运用的战略和强非线性问题,为了维持遗传的多样性,突然变异的概率要适当地设定得大些。

⑤ 探索空间的缩小(Zooming)。GA 法是在离散化设计空间探索。因此,它具有即使探索到解的附近,也很难得到比较精确的最优解的缺点。这里,使用在经过若干世代探索求得的最小解仍没有变化的情况下,以最小解为中心缩小探索空间再进行探索的方法,以提高探索精度。

⑥ 计算结果。GA 的计算参数如表 38.3-2 所示。所求得解用表 38.3-3 表示,其主谐振解用图 38.3-40 表示。相对于世代进化过程,GA 探索的适应度函数最小值变化用图 38.3-41 表示。考察图 38.3-40 可以看出,用 GA 所求的解同近似频响曲线的解比较,具有相当好的精度。

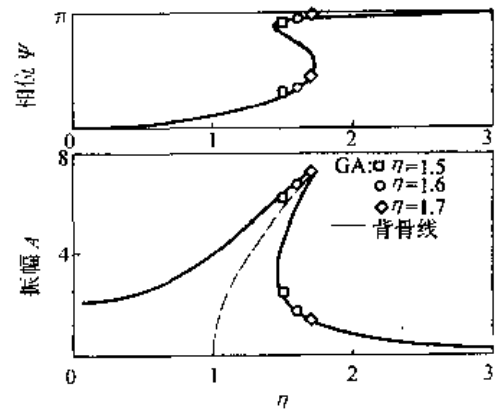


图 38.3-40 GA 的计算结果

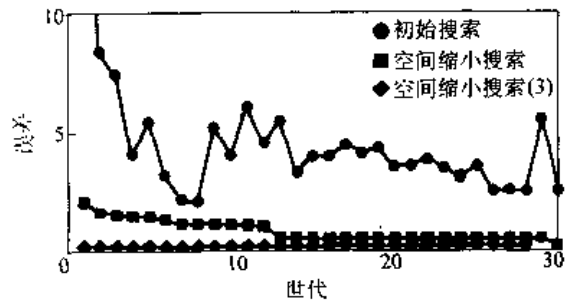


图 38.3-41 GA 的搜索过程中的误差

本例题把 GA 应用于同时求解 Duffing 型非线性

振动共振谐波的复数解问题,讨论了问题的模型、GA的适应度函数和具体的算法。通过计算实例,说明了求解非线性振动解的GA解法。

表 38.3-2 计算参数

个体数(N)	100
染色体长度	8
交叉率(P <sub>c</sub> )	40%
突然变异率(P <sub>m</sub> )	0.10%

表 38.3-3 计算结果

X	η		
	1.5	1.6	1.7
u <sub>1</sub>	-2.3668	1.7324	-1.3875
v <sub>1</sub>	0.7206	0.3294	0.1941
p <sub>1</sub>	-0.0128	0.0019	-0.0137
q <sub>1</sub>	0.0108	0.0039	0.0137
u <sub>2</sub>	3.0118	2.9961	1.1275
v <sub>2</sub>	5.1702	5.9646	7.0513
p <sub>2</sub>	-0.1735	-0.2000	-0.1549
q <sub>2</sub>	0.0422	-0.0353	-0.2137

### 3.5 基于进化的圆抛物面天线健壮结构设计

这里,利用健壮性产品或过程开发方法对一10m圆抛物面天线进行健壮性设计,使天线的性能具有健壮性的设计变量及性能参数值。同时进行多目标优化设计,并将它们的设计结果进行比较。

#### 3.5.1 圆抛物面天线结构设计的要求和特点

在航空航天、卫星通信、雷达技术及射电天文中广泛使用着圆抛物面天线。一般采用的是典型的前馈式反射面天线,其结构为:铺设在背架上的反射面接收来自空间的电磁波,反射会聚到馈源;若是发射,则与之相反。为了对准和跟踪目标,伺服机械带动小齿轮驱动俯仰大齿轮,使天线绕俯仰轴转动以改变仰角α,座架在圆形轨道上转动以改变方位。天线结构一般指可俯仰转动的部分,它常常被简化为桁架结构。一般天线结构的载荷主要是自重、风力、温度、冰雪和冲击振动。对大型高精度天线,由于有天线罩和较好的工作环境,其载荷主要是自重。天线在近些年的研究中获得了广泛的重视,是因为天线是一种高精度的结构(如图38.3-42所示的10m圆抛物面天线)。其设计要求如下:

##### (1) 天线反射面的精度

天线结构与一般结构不同,有其特殊的要求,天线反射面的变形误差会影响到电性能。严格地说,反射面误差对电性能的影响应当按照电磁场理论,把反射面作为电磁场的边界条件来处理。但实际设计中采用近似的几何光学原理和以统计规律为基础的鲁兹

(Ruze)公式来估算这种影响仍有一定的参考价值;由于表面误差引起了电磁波的光程差,使口面上不是等相位面,而造成天线增益的下降等影响。这种影响可用鲁兹公式来估算:

$$\eta_r = \frac{G}{G_0} = e^{-4(\pi\delta/\lambda)^2} \quad (38.3-55)$$

式中 η<sub>r</sub>——增益下降系数(天线表面效率);

G<sub>0</sub>——无表面误差时的增益;

G——有表面误差时的增益;

δ——表面各点半光程差的均方差值(RMS);

λ——波长。

由上式可见,随着表面误差的增大,天线增益急剧下降,当δ=λ/30时,η<sub>r</sub>=83.9%;而当δ=λ/16时,η<sub>r</sub>=54.1%,后者意味着误差为δ=λ/16时,天线口面只能相当于无误差天线的一半面积,因此,这是最低极限。工作在米波、厘米波及毫米波的天线,根据工程要求,其精度指标往往取为λ/16~λ/60。由此可见,与一般结构相比,天线结构的精度要求是非常高的。

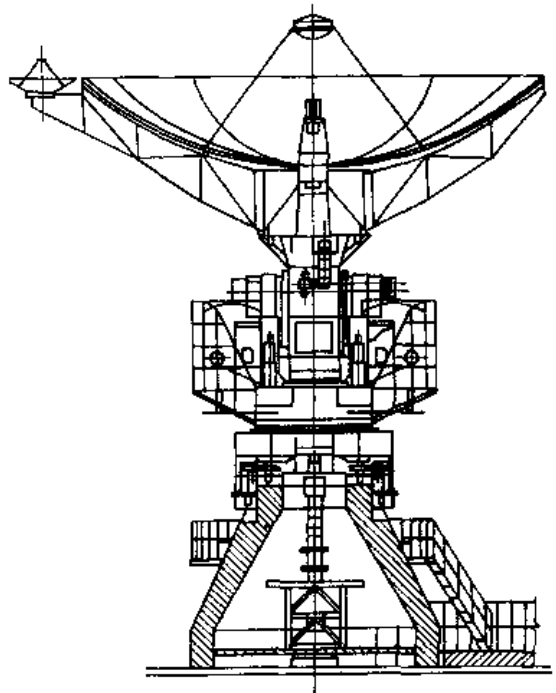


图 38.3-42 10m 圆抛物面天线整体结构图

##### (2) 天线结构的自重变形要求

对于工作波长较短、表面精度要求较高的天线,其变形必须有严格的要求,这就决定了对天线结构的要求不同于一般的结构。它首先要求满足精度、刚度要求,强度条件则往往不成问题。例如,德国的某直径为28.5m的天线,在风力55.6m/s、3cm冰厚及自重等载荷作用下,最大应力约为0.03MPa,远小于许用应力。

科学技术的发展需要工作于厘米波、毫米波段的增益很高的天线,也就是说,要求天线尺寸尽可能大,精度又很高,这就出现了天线结构精度与自重变形的尖锐矛盾。大型高精度反射面天线由于要求刚度好,所以自重较大,自重载荷成了大型精密天线的主要载荷。

(3) 最佳吻合抛物面与保型要求

影响天线电性能的并非表面点位移的绝对数值,而是反射表面自身的相对变形。1967年,冯·霍纳(Von Hoerner)提出了保型设计的思想。他设想:如能设计出一种天线结构,变形后反射表面相对最佳吻合抛物面的误差为零,这就是保型设计。也就是说,天线变形后,反射表面仍然是同族的反射曲面——抛物面。当天线从一仰角转到另一仰角时,反射面由一抛物面变成另一抛物面,这就是“保型变形”。要使天线在任何仰角上反射面都是抛物面,只要使天线在仰天和指平两位置变形后仍为理想抛物面即可,这是因为在任意仰角位置上的天线,在自重作用下的变形,可以表示为天线在仰天和指平两位置自重变形的线性组合。

严格的保型设计并非易事。由于理论上严格保型设计比较困难,工程上由于制造、安装及其他困难,严格保型也难以达到,所以目前的天线保型设计大多数是近似保型设计,即对天线结构进行优化设计,使天线反射面变形后,相对其最佳吻合面半光程差的均方根减小到工程设计提出的指标范围内。

(4) 天线结构优化设计的特点

对天线结构优化设计的要求主要是:表面精度高,结构质量小,转动惯量小,在各种环境载荷下不破坏,谐振频率高,容易加工安装,造价低廉。不同的天线类型对以上要求侧重不同,优化设计提法也不同。天线结构优化设计不同于一般结构优化设计的特点是:

- 1) 刚度大,精度要求高,这是天线优化设计的主要矛盾。
- 2) 自重载荷是结构的主要载荷,优化中设计变量变化引起的自重载荷变化的导数不容忽视。
- 3) 结构为高次静不定的大型结构。
- 4) 与一般结构个别点位移约束相应的是反射面所有点的半光程差均方根值精度约束或目标函数。

5) 工程设计要求决定了优化设计的目标多、约束多。各目标与约束之间没有明确的区别,即目标函数和约束条件可以相互转化,要根据具体工程要求和优化中结构设计方案的实际状态确定目标函数和约束条件的选取。

3.5.2 天线反射面精度计算

天线反射面的精度严重地影响着天线的电性能。

公式(38.3-55)表明,与天线增益有直接关系的是表面各点半光程差的均方根值。本节将推出圆抛物面天线结构变形后表面点相对原设计抛物面以及最佳吻合抛物面半光程差的计算公式,并给出最佳吻合抛物面吻合参数的求法。

(1) 光程差

根据抛物面的性质,位于焦点  $F$  的馈源向抛物面发出的电磁波,经反射后成为平行于轴线  $FO$  的射线,且各条射线到达垂直于轴线的平面的路程相等,即  $FA + AA' = FB + BB'$  (见图 38.3-43)。

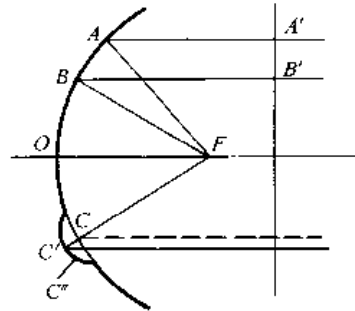


图 38.3-43 抛物面的光程差

(2) 表面点位移引起的对原设计面的半光程差

设  $OXYZ$  为原设计抛物面的坐标系,原点  $O$  为抛物面顶点, $OZ$  为抛物面焦轴, $f$  为焦距, $x, y, z$  为表面点的坐标,它们满足如下设计抛物面方程

$$x^2 + y^2 = 4fz \quad (38.3-56)$$

下面推导表面点各向位移  $u, v, w$  引起的半光程差(图 38.3-44)。

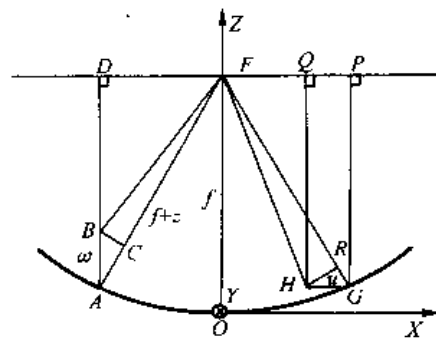


图 38.3-44 各向位移的半光程差

1) 轴向位移引起的半光程差 如图 38.3-44 左所示,如果表面点  $A$  由于轴向位移  $w$  移至  $B$  点,则从焦点  $F$  到口面的光程长度从  $FA + AD = 2f$  变为  $FB + BD$ ,注意到  $w$  是小变形量,如取  $FC = FB$ ,必有  $BC \perp AF$ ,于是两光程之差为

$$N = (FA + AD) - (FB + BD) = AB + AC = w(1 + AC/AB) = w(1 + AD/AF)$$

而

$$AF = \sqrt{(f-z)^2 + x^2 + y^2}$$

$$\sqrt{(f-z)^2 + 4f^2} = f+z$$

于是  $\Delta' = w \left( 1 + \frac{f-z}{f+z} \right) = w \frac{2f}{f+z}$

以光程增加为正,则位移  $w$  引起的半光程差为

$$\Delta_1 = \frac{fw}{f+z} \quad (38.3-57a)$$

2) 位移  $u, v$  引起的半光程差 如图 38.3-44 右方所示,如果表面点  $G$  由于  $z$  向位移  $u$  移至  $H$  点,则从焦点  $F$  到  $\Pi$  面的光程长度从  $FG+GP$  变为  $FH+HQ$ ,注意到  $u$  是小变形量,如取  $FR=FG$ ,必有  $HR \perp FG$ ,于是,两光程之差为

$$\begin{aligned} \Delta'' &= (FG+GP) - (FH+HQ) = GR \\ &= u \cdot GR/HG - u \cdot FP/FG = ux/(f+z) \end{aligned} \quad (38.3-57b)$$

同理,可得位移  $v$  的光程差为

$$uy/(f+z)$$

于是由位移  $u, v$  引起的半光程差为

$$\Delta_2 = \frac{xu + yv}{2(f+z)} \quad (38.3-58)$$

3) 全部半光程差 表面点各向位移  $u, v, w$  引起的全部半光程差为

$$\Delta = \Delta_1 + \Delta_2 = \frac{xu + yv - 2fw}{2(f+z)} \quad (38.3-59)$$

### 3.5.3 最佳吻合抛物面各点对原设计面相应点的半光程差

#### (1) 最佳吻合抛物面及其吻合参数

对于圆抛物面天线,设原设计面为  $A$ ,变形后的反射曲面为  $B$ 。对于  $B$ ,总可以找到一个最佳吻合抛物面  $BFP$  图(38.3-45)。在原设计面的坐标系  $OXYZ$  中,设  $BFP$  对  $A$  的顶点位移为  $u_A, v_A, w_A$ ,按右手螺旋定向的轴线转角为  $\varphi_x, \varphi_y$ ,焦距  $f$  的增量为  $h, BFP$  具有其相应的坐标系为  $O_1X_1Y_1Z_1$ 。

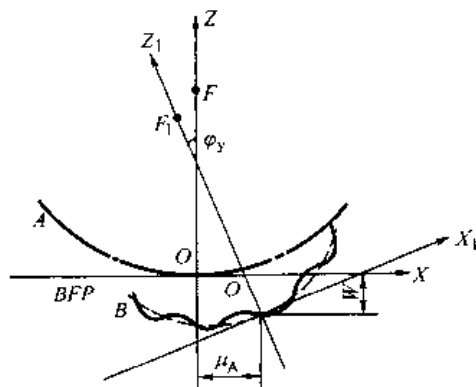


图 38.3-45 最佳吻合抛物面

#### (2) 最佳吻合抛物面各点对原设计面相应点的位移

考虑到反射面变形位移及其吻合参数均为微量,

可以忽略其二阶微量,由图 38.3-46 可求得最佳吻合抛物面上各点对原设计面相应点的位移为

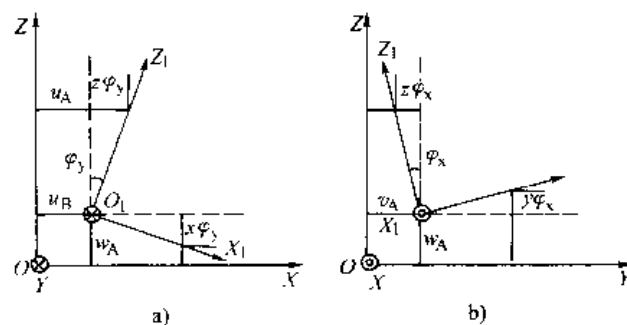


图 38.3-46 最佳吻合抛物面各点的位移

$$\left. \begin{aligned} u' &= u_A + z\varphi_y \\ v' &= v_A - z\varphi_x \\ w' &= w_A - x\varphi_y + y\varphi_x - hz/f \end{aligned} \right\} \quad (38.3-60)$$

其中,  $-hz/f$  项是由焦距增量引起的。这是因为,由  $x^2 + y^2 = 4fz$  可得  $z = (x^2 + y^2)/(4f)$ ,由焦距增量  $h$  引起的  $z$  向位移可表示为如下微分形式(图 38.3-47):

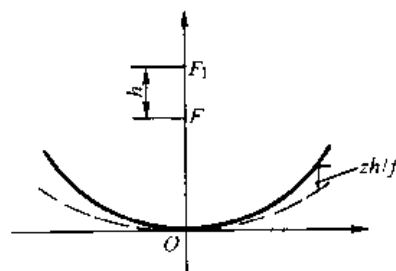


图 38.3-47 焦距增量引起的位移

$$dz = \frac{dz}{df} h = \frac{x^2 + y^2}{4} \cdot \left( \frac{-1}{f^2} \right) \cdot h = -\frac{4fzh}{4f^2} = -\frac{zh}{f}$$

#### (3) 最佳吻合抛物面对原设计面相应点的半光程差

将式(38.3-60)代入式(38.3-59),即可得最佳吻合抛物面对原设计面相应点的半光程差为

$$\begin{aligned} \Delta &= (xu' + yv' - 2fw')/2(f+z) \\ &= \frac{1}{2(f+z)} [xu_A + yv_A \\ &\quad - 2fw_A - 2hz - y(x+2f)\varphi_x + \\ &\quad x(x+2f)\varphi_y] \end{aligned} \quad (38.3-61)$$

#### (4) 表面点位移对最佳吻合抛物面的半光程差

从表面点位移对原设计的半光程中,减去最佳吻合抛物面相应点对原设计面的半光程差,就是表面点  $i$  对最佳吻合抛物面的半光程差,即以式(38.3-59)减式(38.3-61),可得

$$\begin{aligned} \delta_i &= \frac{1}{2(f+z_i)} [x_i(u_i - u_A) + y_i(v_i - v_A) \\ &\quad - 2f(w_i - w_A) - 2hz_i + y_i(z_i + 2f)\varphi_x \\ &\quad - x_i(z_i + 2f)\varphi_y] \end{aligned} \quad (38.3-62)$$

(5) 表面点位移对最佳吻合抛物面的加权半光程差均方根

引入各表面点的加权因子

$$\rho_i = (n_0 q_i a_i) / \left( \sum_{i=1}^{n_0} q_i a_i \right) \quad (38.3-63)$$

其中,  $a_i$  为表面点  $i$  影响的反射面积;  $q_i$  为照度因子, 有

$$q_i = 1 - C r_i^2 / R_0^2 \quad (38.3-64)$$

其中,  $r_i$  为表面点  $i$  到焦距的距离;  $R_0$  为口面半径;  $n_0$  为表面节点总数;  $C$  为由焦距比  $f/R_0$  决定的常数。

于是, 可得表面所有节点位移引起的对最佳吻合抛物面的加权半光程差均方根值, 为

$$\delta = \left( \sum_{i=1}^{n_0} \frac{\rho_i \delta_i^2}{n_0} \right)^{\frac{1}{2}} \quad (38.3-65)$$

这就是直接影响天线电性能的反射面精度指标。为了研究方便, 这里以加权因子相同处理。

#### 3.5.4 10m 圆抛物面天线健壮设计模型

(1) 10m 圆抛物面天线的结构形式、特点

10m 圆抛物面天线, 其直径  $D=10000\text{mm}$ , 焦距  $f=3000\text{mm}$ 。采用卡塞格伦抛物面, 方程为

$$y = \frac{x^2 + z^2}{4f} \quad (38.3-66)$$

随着天线探测目标飞行速度的加快, 要求雷达天线能在极短的时间内精确地测出目标的方位、仰角和斜度。研究的天线反射体结构形式为典型的辐射状桁架结构的主力骨架。在结构设计中需要考虑的是:

- 1) 中心结构的形式。
- 2) 辐射梁数与环数。
- 3) 梁高。
- 4) 梁的腹杆布置。
- 5) 对角斜杆。
- 6) 反射面板。

大型天线的面板一般都是分块的。从安装方便这一点来考虑, 希望分块面板的尺寸大而块数少。但是从保证面板制造精度、减少热变形来考虑分块面板的尺寸应该小些。反射体主力骨架制造不易准确, 因此反射面板安装不能以主力骨架为基准, 而是用调节螺杆将反射面板连接到主力骨架上, 反射面板与骨架间的距离可以调节。通过螺杆的调整使反射面板精确地定位。对高精度的反射面板, 为增加刚性面又使重量轻, 常采用铝型材与铝板铆接在一起组成加强肋结构。在设计时, 径向加强肋可以多些, 环向加强肋尽量少些, 因为环向加强肋加工困难。本天线的面板采用 2mm 铝板。

(2) 根据设计要求建立的力学模型及数学模型

设计功能参数为:

① 要求  $G \leq 4t$ 。

② 变形后天线表明各节点位移的均方根值  $\leq 0.7\text{mm}$ 。

③ 8 级风  $60^\circ$  仰角时正常跟踪, 12 级风朝天时不受损坏。

④ 要求机械谐振频率在 12Hz 以上。

根据设计要求, 建立了天线体的有限元模型。整个天线体的有限元模型如图 38.3-48 所示。模型的详细情况:

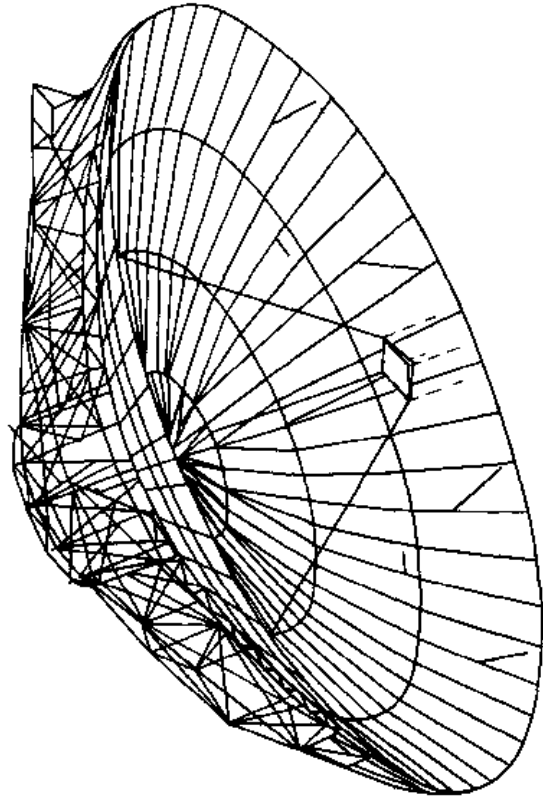


图 38.3-48 10m 圆抛物面天线反射体模型

1) 模型共有 648 个节点, 分为三种单元: 杆单元, 梁单元, 板壳单元。杆单元数量为 168 个, 梁单元数量为 763 个, 板壳单元数量为 206 个。天线的面板由 12 块厚度为 2mm 的铝板组成, 互相之间没有连接, 只连接在骨架定位。另外为了增加刚度, 在大线背面有若干铝制加强肋, 分为径向加强肋和环向加强肋, 每隔  $7.5^\circ$  有一条径向的加强肋。边框的加强肋采用槽铝, 中间采用 Z 形铝, 这样便于铆接。在辐射梁和支承梁之间有相应数量的连杆固定, 连杆及对角斜杆采用杆单元。杆单元有二组截面形式, 梁单元有 16 组截面形式。

2) 设计变量 建立圆抛物面天线的数学模型时, 以图 38.3-49 中主梁标号 (1)~(10) 的起主要承载作用的 10 根梁的截面特性作为设计变量, 因为这 10 根梁可能对整个天线体的性能有较大的影响, 而且在天线体中共有 12 组这样的辐射梁, 如果能尽量减少这些梁的重量而又使整个天线体的精度仍足够, 同时仍能满

足其他各种性能的要求,如基频、强度、刚度的要求,那么对天线的设计和制造来说,其成本将会有很大的降低;同时,进一步考虑到健壮性的要求,使得当天线的主要承载部件发生变异时,其所有性能仍能在一定范围内保持不变,那将会带来三个方面的好处:

① 一是在初始设计阶段给关键设计变量一个较大的调整范围而同时保持天线性能,在详细设计阶段,允许对某些变量做较大的改变来适应设计的需要,具有可调整性(Modifiable),同时又无需重新对整个结构作出调整和验算,使设计具有柔性(Flexibility)。

② 另一个好处是在制造过程中允许更大范围的制造误差而仍不算次品,这样提高了成品率,即提高了生产率(Productivity)。

③ 还有一个好处是在使用过程中即使由于内部条件(如部件老化等)及外部条件(如环境温度等)的影响使承载部件发生改变,但仍使得整体的性能不变或仅有很小的变化,使成品的质量得以保证,具有健壮性(Robustness)。

3) 建立多目标优化模型(MOP) 以天线的重量  $G$  和天线反射表面的精度  $\delta$  作为设计目标,以天线体的基频  $f_0$  和强度、刚度作为约束条件,建立如下的多目标优化模型:

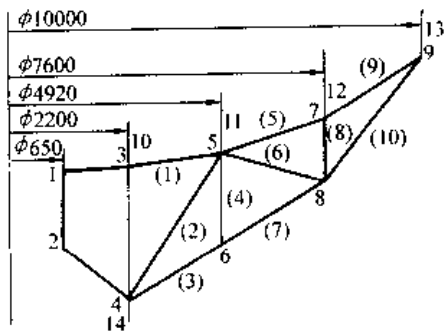


图 38.3-49 主梁的腹杆布置

$$\begin{aligned} \min f(1) &= G \\ \min f(2) &= \delta \end{aligned} \quad (38.3-67)$$

$$\begin{aligned} \text{s. t. } g(1) &= 12/f_0 - 1 \leq 0 \\ g(2) &= \sigma_{\max}/\sigma_s - 1 \leq 0 \\ g(3) &= f(1)/0.7 - 1 \leq 0 \end{aligned}$$

4) 加入健壮性设计 在建立的多目标优化模型的基础上,将加入健壮性的考虑,即将质量加入到设计过程中,形成一个新的健壮性设计的数学模型。对天线体的健壮性设计仅考虑可控因素的变异造成系统性能变异的情况,即第二类健壮性设计。

### 3.5.5 10m 圆抛物面天线体结构的健壮性设计过程

第1步,确定设计变量的个数,取值范围

从上面的分析来看,这里初步以图 38.3-49 中主梁标号(1)~(10)的起主要承载作用的 10 根梁的截面特性作为设计变量。10 根梁的截面特性为:

梁(1)(2)(5)(9)(10)取为  $\phi 50\text{mm}$ ,厚度从 2mm 到 4mm,3mm 厚度时取为 0 水平,2mm 和 4mm 时分别取为 -1 和 1 水平;

梁(4)(6)(8)取为  $\phi 40\text{mm}$ ,厚度从 1mm 到 3mm,2mm 厚度时取为 0 水平,1mm 和 3mm 时分别取为 -1 和 1 水平;

梁(3)(7)取为  $\phi 65\text{mm}$ ,厚度从 3mm 到 5mm,4mm 厚度时取为 0 水平,3mm 和 5mm 时分别取为 -1 和 1 水平;

冷弯无缝钢管的标准尺寸见表 38.3-4。

这样,10 根梁按三水平取值的范围见表 38.3-5。

表 38.3-4 冷拔无缝钢管的标准尺寸

钢管外径	可选壁厚							
	1.0	1.2	1.4	1.5	1.6	1.8	2.0	
$\phi 40$			2.2	2.5	2.8	3.0		
$\phi 50$	2.0	2.2	2.5	2.8	3.0	3.2	3.5	4.0
$\phi 65$		3.0	3.2	3.5	4.0	4.5	5.0	

表 38.3-5 10 根梁的三水平取值范围

梁号	水平		
	-1	0	1
1	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
2	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
3	$\phi 65 \times 3$	$\phi 65 \times 4$	$\phi 65 \times 5$
4	$\phi 40 \times 1$	$\phi 40 \times 2$	$\phi 40 \times 3$
5	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
6	$\phi 40 \times 1$	$\phi 40 \times 2$	$\phi 40 \times 3$
7	$\phi 65 \times 3$	$\phi 65 \times 4$	$\phi 65 \times 5$
8	$\phi 40 \times 1$	$\phi 40 \times 2$	$\phi 40 \times 3$
9	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
10	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$

第2步,研究单个因素对系统性能的影响程度

本步的目的为研究在其他因素值固定的情况下,某一因素的水平变化时,对系统性能的影响,从而确定单个因素对系统的贡献。方法为选某一因素,其水平在 -1,0,1 三种状态而其他因素固定在 0 水平时系统响应的变化。系统响应有 4 个,即天线体的重量  $G$ ,天线反射表面的精度  $\delta$ ,天线体的基频  $f_0$ ,以及在承载时天线体中的最大应力状况  $\sigma_{\max}$ 。

经过计算后得到 10 组数据,用图形的方式显示出来。图中横坐标为某一因素的三个水平,纵坐标为正规化后的系统响应,分别为



$$G' = G/G_{max} \quad (38.3-68)$$

$$\delta' = \delta/\delta_{max} \quad (38.3-69)$$

$$f'_0 = f_0/f_{0max} \quad (38.3-70)$$

$$\sigma'_{max} = \sigma_{max}/\sigma_{max0} \quad (38.3-71)$$

以上四式中  $G_{max}$  为允许最大重量 4t;  $\delta_{max}$  为允许最大天线反射表面的误差 0.7mm;  $f_{0max}$  为 10 个设计变量都取 1 水平时的基频 40.347Hz;  $\sigma_{max0}$  为钢材的屈服点值 235MPa。因得到的最大应力值均在 50MPa 左右, 离屈服点值太远, 偏于安全, 在图中不表示出来。

图 38.3-50~图 38.3-63 是为了研究每一个响应在系统各因素的水平发生变化时对响应的贡献程度, 以便比较每个因素对系统响应的贡献。从图中可以看出:

1) 上面计算的各种情况下, 最大应力都出现在钢制主力结构上, 铝制部件由于不起主要的承载作用, 应力很小, 并且计算得到的最大应力远远小于屈服点。由此可知, 在本天线的设计中, 强度约束不是起作用约束。

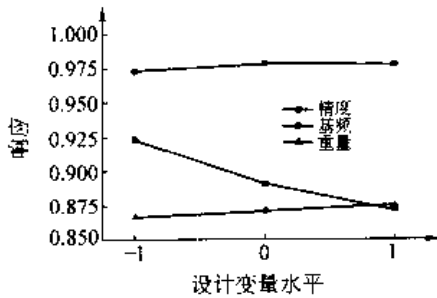


图 38.3-50 设计变量(1)变化引起响应的变化

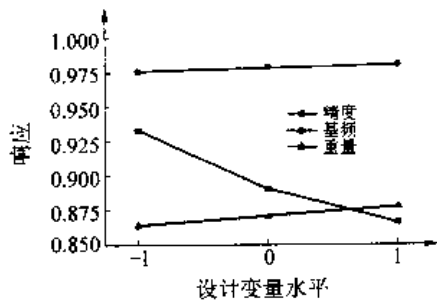


图 38.3-51 设计变量(2)变化引起响应的变化

2) 因素(1)~因素(8)的基本规律都是当因素的水平增加时, 天线体的重量增加, 基频也增加, 同时天线反射表面点位移对最佳吻合抛物面的加权半光程差均方根值减小, 即天线的精度增加。其中的例外是基频对因素(1)的响应出现了弯曲(图 38.3-50), 由于数据接近, 图中并不一定看得清楚, 还有反射面精度对因素(4)的响应也出现了弯曲(图 38.3-53)。

3) 因素(9)、因素(10)造成反射面精度的响应出现弯曲。(图 38.3-58、图 38.3-59)

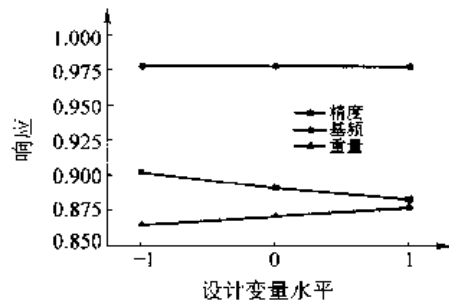


图 38.3-52 设计变量(3)变化引起响应的变化

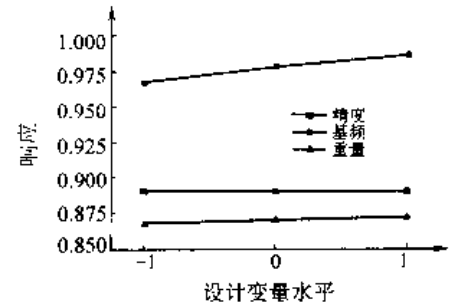


图 38.3-53 设计变量(4)变化引起响应的变化

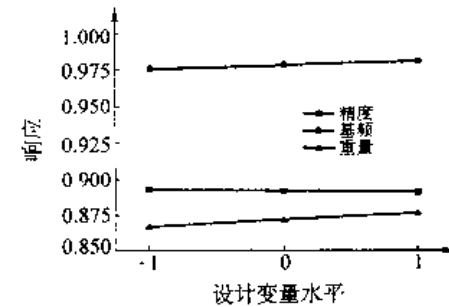


图 38.3-54 设计变量(5)变化引起响应的变化

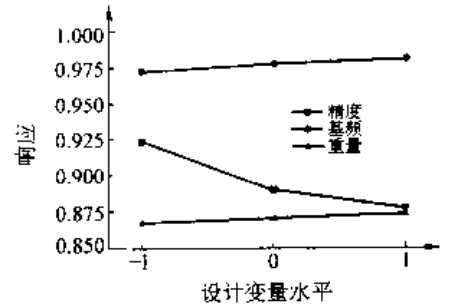


图 38.3-55 设计变量(6)变化引起响应的变化

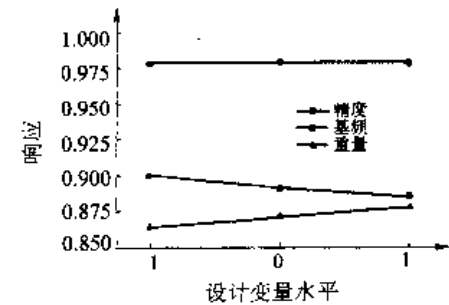


图 38.3-56 设计变量(7)变化引起响应的变化

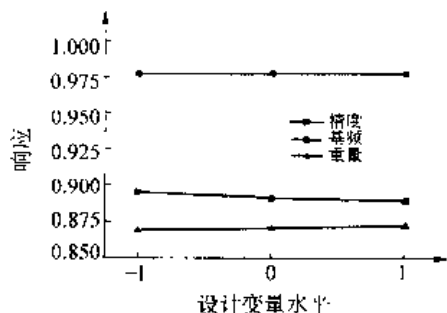


图 38.3-57 设计变量(8)变化引起响应的变化

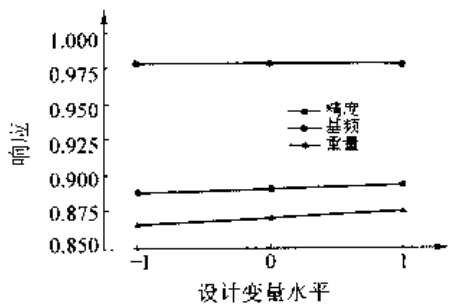


图 38.3-58 设计变量(9)变化引起响应的变化

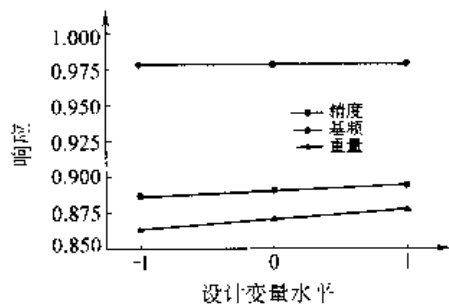


图 38.3-59 设计变量(10)变化引起响应的变化

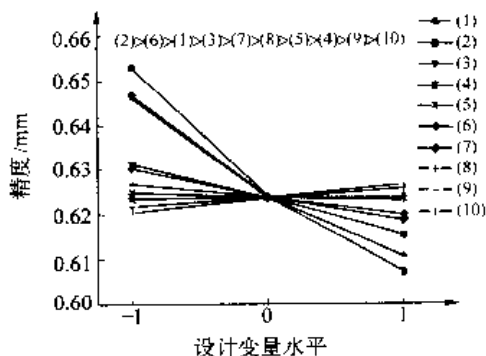


图 38.3-60 各因素的水平引起反射面精度的变化

4) 对精度影响最大的因素是(2)、(6)、(1), 其余因素的影响不大。影响程度从大到小为(2)→(6)→(1)→(3)→(7)→(8)→(5)→(4)→(9)→(10)。因素(4)的变化对精度几乎没有影响。一般的规律是随着各因素水平的增加, 反射面精度也增加, 但因素(9)和(10)正好相反, 当其水平数增加时, 反射面精度反而降低。分析原因, 是因为梁(9)和(10)的钢管厚度增加后, 引起

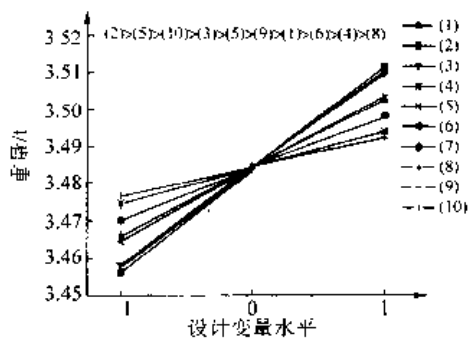


图 38.3-61 各因素的水平引起天线体重量的变化

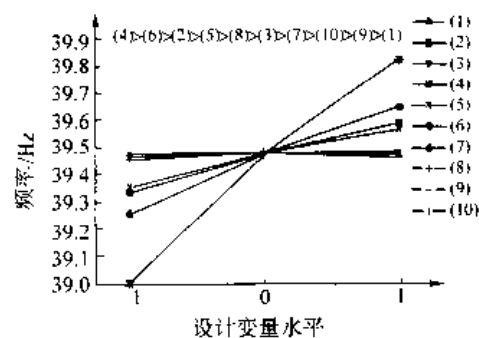


图 38.3-62 各因素的水平引起天线反射体基频的变化

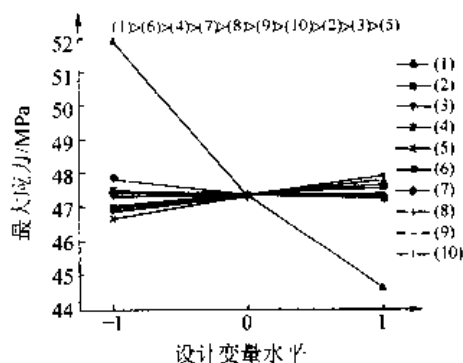


图 38.3-63 各因素的水平引起天线体最大应力的变化  
局部刚性过大, 反射面整体变形不均匀, 使表面点对最佳吻合抛物面的加权半光程差的均方根值反而增大。由此得到的结论是组成辐射梁的各梁截面变化应比较均匀, 使整体受力、变形的过渡平缓。

5) 天线体重量随着因素水平的增大而增加。各因素对重量影响程度从大到小的次序为: (2)→(5)→(10)→(3)→(5)→(9)→(1)→(6)→(4)→(8)。

6) 天线体的基频随着各因素水平的增加而增大, 唯一的例外是因素(1), 出现了弯曲, 呈非线性规律。各因素的影响程度从大到小为: (4)→(6)→(2)→(5)→(8)→(3)→(7)→(10)→(9)→(1)。

7) 对天线体的最大应力影响程度从大到小的次序为(1)→(6)→(4)→(7)→(8)→(9)→(10)→(2)→(3)→(5)(见图 38.3-63), 基本规律是各因素的水平增加时, 天线体中的最大应力减小, 但其中因素(9)、(10)

的水平增加时,天线体中的最大应力反而增大;而因素(3)、(5)的水平增加造成了最大应力响应的弯曲。

8) 从图中可以看出,天线的反射面精度和重量是一对相互冲突的目标,如果要减轻重量,一般会导致反射面精度的下降;反之,反射面精度的提高(即反射表面点对最佳吻合抛物面的加权半光程差的均方根值减小)一般会导致天线重量的增大,所以要寻求各参数的最佳搭配,使两个目标都最大限度地接近最小值。

9) 因素(9)、(10)的水平增加会导致反射面精度和重量以及最大应力同时增大,基频的变化很小,由此可以将因素(9)、(10)固定在一1水平,以后作为常数

使用。

10) 因素(4)的水平增加会导致基频和重量增大,而反射面精度几乎不变。基频对因素(4)的响应虽然受影响较大,但基频的绝对变化并不大,远高于要求的最小基频。所以将因素(4)固定在一1水平,作为常数使用。

11) 因素(5)、(8)对重量和精度的影响都很小,将这两个因素固定在0水平,作为常数使用。

为了进一步研究天线辐射梁的参数对天线性能的影响,又取了代表梁高的参数,即6、8两点的纵坐标 $y_A$ 和 $y_B$ 作为设计变量,所取的水平与其代表的实际值分别如表38.3-6所示。

表 38.3-6 梁高设计变量取值范围

	设计变量所取的水平数与其代表的实际值/mm				
	-2	-1	0	1	2
$y_A$	1110	910	710	510	310
对应梁(4)高	535	735	935	1135	1335
$y_B$	1990	1790	1590	1390	1190
对应梁(8)高	350	350	750	950	1150

图 38.3-64 中绘出了  $y_A$  和  $y_B$  单独变化同时其余因素取 0 水平时的响应变化趋势。

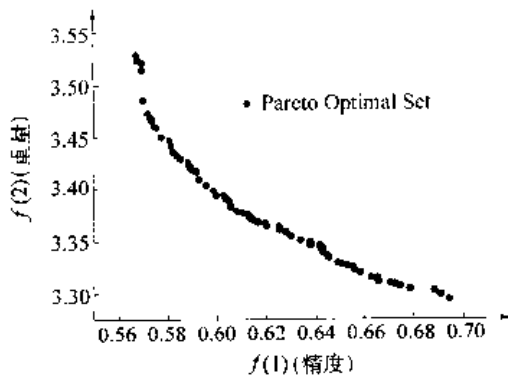


图 38.3-64 天线多目标优化的 Pareto 最优解集

从图中可看出,梁高  $y_A$  和  $y_B$  增加时,随之而来的是天线体重量的增加,是必然的规律,同时天线反射面表面精度呈二次曲线状态, $y_A$  和  $y_B$  处于 0 水平时精度最高; $y_A$  的增加导致基频的降低,而  $y_B$  的增加导致基频的增加。由结构中的最大应力状况看出(数据未列出), $y_A$  的增加导致最大应力增加,而  $y_B$  的增加导致最大应力减小。图中的变化趋势表明, $y_A$  对系统响应的影响较大, $y_B$  的影响较小。

结论:通过上面的分析,发现有的因素对系统的影响很小,或改变该因素会使系统的目标同时变好或变坏,对这些因素的处理是寻求其水平范围内的最佳点固定下来作为常数处理,而其余因素作为主要影响因素进入下一步的研究。固定下来的因素为:(4)、(5)、(8)、(9)、(10)。

还要注意的,本步骤研究的是单个因素对系统

的影响,即假设各因素独立,因素间无交互作用。事实上,下一步的研究中将看到系统的各因素间可能存在的交互作用。

第 3 步,规划序列试验设计,并拟合响应面方程

将上一步预选出的主要因素整理成新的变量。因本例题研究的是可控因素的健壮性设计,用  $x_1, x_2, x_3, x_4, x_5, x_6, x_7$  表示因素(1)、(2)、(3)、(6)、(7)及  $y_A$  和  $y_B$ 。各因素取 3 水平,各水平代表的实际值如表 38.3-7 所示。

表 38.3-7 主要因素的水平及其实际值 (mm)

因素	水平		
	-1	0	1
$x_1$ ——(1)	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
$x_2$ ——(2)	$\phi 50 \times 2$	$\phi 50 \times 3$	$\phi 50 \times 4$
$x_3$ ——(3)	$\phi 65 \times 3$	$\phi 65 \times 4$	$\phi 65 \times 5$
$x_4$ ——(6)	$\phi 40 \times 1$	$\phi 40 \times 2$	$\phi 40 \times 3$
$x_5$ ——(7)	$\phi 65 \times 3$	$\phi 65 \times 4$	$\phi 65 \times 5$
$x_6$ —— $y_A$	1110	710	310
$x_7$ —— $y_B$	1990	1590	1190

主要因素中  $x_1 \sim x_7$  为离散变量,其正规化后的取值范围为

$$x_1 \text{ 和 } x_2: [-1, -0.8, -0.5, -0.2, 0, 0.2, 0.5, 1]$$

$$x_3 \text{ 和 } x_5: [-1, -0.8, -0.5, 0, 0.5, 1]$$

$$x_4: [-1, -0.8, -0.6, -0.5, -0.4, -0.2, 0, 0.2, 0.5, 0.8, 1]$$

$x_6, x_7$  为连续变量,其正规化后的取值在  $[-1, 1]$  之间,与原变量之间的对应关系为

$$x_1 = \frac{71 - y_A}{40} \quad (38.3-72)$$

$$x_2 = \frac{159 - y_B}{40} \quad (38.3-73)$$

三个响应分别为  $y_1$  (MSE 均方误差)、 $y_2$  (重量)、 $y_3$  (基频)。

先规划低阶试验设计,拟合一阶响应面方程。有 7 个因素的一阶响应面方程为

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

$$- b_5x_5 + b_6x_6 + b_7x_7 \quad (38.3-74)$$

方程中有 8 个待定系数,需要至少 8 次试验值来拟合。拟合数据恰好等于 8 组的试验设计叫饱和试验设计,前面已分析过,饱和试验设计产生的结果并不理想,因此需要至少多于待定系数个数的试验次数。这里选用  $2^{k-p}$  分式析因设计,即  $2^{7-2}$  分式析因设计,共 32 次试验。其响应都是可计算的,所以采用计算机仿真,由有限元程序计算即可得到。试验安排及计算所得三个响应的值如表 38.3-8 所示。

表 38.3-8  $2^{7-2}$  分式析因设计及计算结果表

试验号	因素及所取的水平							响应		
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$y_1/\text{mm}$	$y_2/t$	$y_3/\text{Hz}$
1	-1	1	-1	1	1	-1	-1	0.6863	3.407	39.1114
2	1	-1	-1	1	-1	-1	1	0.7014	3.135	40.8131
3	1	-1	-1	-1	-1	-1	-1	0.7419	3.3041	39.0069
4	1	1	1	1	1	-1	-1	0.6327	3.5037	39.5324
5	-1	-1	-1	1	1	-1	1	0.7382	3.3947	39.3659
6	1	1	1	-1	-1	-1	-1	0.7059	3.4198	39.2807
7	-1	1	-1	-1	1	-1	1	0.8161	3.4205	39.0153
8	1	1	1	-1	-1	1	1	0.632	3.4686	38.8504
9	-1	1	1	1	-1	1	-1	0.6931	3.4009	38.1615
10	-1	-1	1	1	1	-1	1	0.6720	3.4107	38.8601
11	1	1	-1	1	1	1	-1	0.6532	3.493	38.9171
12	1	-1	1	-1	-1	1	-1	0.7488	3.378	38.3324
13	-1	1	1	-1	-1	-1	1	0.8101	3.4346	39.0295
14	-1	-1	1	1	1	1	1	0.7304	3.4603	38.7196
15	-1	1	-1	1	-1	-1	1	0.7233	3.4051	39.5968
16	1	1	-1	-1	1	1	1	0.6324	3.4757	38.8461
17	1	-1	1	-1	1	1	1	0.7153	3.4671	38.5469
18	-1	-1	1	1	-1	-1	1	0.7283	3.4088	39.3738
19	-1	-1	1	-1	-1	-1	-1	0.7436	3.3268	38.5582
20	1	-1	-1	1	1	-1	-1	0.6685	3.388	39.2668
21	-1	1	-1	1	-1	1	-1	0.6931	3.383	38.4685
22	-1	-1	-1	1	-1	-1	-1	0.7204	3.2966	38.8398
23	1	-1	1	-1	1	-1	-1	0.6953	3.4183	39.0247
24	1	-1	1	1	1	1	-1	0.7374	3.4845	38.6102
25	-1	1	-1	-1	1	1	-1	0.6920	3.4265	38.1614
26	1	1	1	1	1	1	1	0.6044	3.5533	39.4035
27	-1	1	1	-1	1	1	1	0.6633	3.486	38.3901
28	-1	-1	1	-1	-1	1	1	0.7526	3.3757	38.0713
29	-1	1	1	1	-1	1	1	0.6460	3.4618	39.0021
30	1	-1	-1	-1	-1	1	1	0.7256	3.3653	38.5344
31	-1	-1	-1	-1	1	1	1	0.7554	3.3827	38.0675
32	-1	-1	-1	-1	-1	1	-1	0.7847	3.2976	37.8316

由表中的数据拟合一阶响应面方程,并进行显著性检验,可以得到如下的结果:

用线性响应面模型拟合系统的响应  $y_1$  (MSE 均方

误差)、 $y_2$  (重量)、 $y_3$  (基频)、 $y_4$  (结构中最大应力),模型的 ANOVA (方差分析) 结果如表 38.3-9~表 38.3-12 所示。

表 38.3-9 响应  $y_1$ (MSE 均方误差)线性模型的方差分析

模型		平方和	df	均方	F	Sig.
1	回归	4.691E-02	7	6.701E-03	4.756	0.002
	残差	3.381E-02	21	1.409E-03		
	求和	8.072E-02	31			

表 38.3-10 响应  $y_2$ (重量)线性模型的方差分析

模型		平方和	df	均方	F	Sig.
1	回归	0.109	7	1.564E-02	78.728	0.000
	残差	4.768E-03	24	1.987E-04		
	求和	0.114	31			

表 38.3-11 响应  $y_3$ (基频)线性模型的方差分析

模型		平方和	df	均方	F	Sig.
1	回归	9.456	7	1.351	31.945	0.000
	残差	1.015	24	4.229E-02		
	求和	10.471	31			

表 38.3-12 响应  $y_4$ (结构中最大应力)线性模型的方差分析

模型		平方和	df	均方	F	Sig.
1	回归	410.330	7	58.619	16.764	0.000
	残差	83.919	24	3.497		
	求和	494.250	31			

从上面的四个 ANOVA 结果来看,四个线性响应面模型的显著性都较高,都可以作为转换模型使用。为了进一步提高精度,可采用可旋转的中心复合设计,在线性模型的基础上追加样本点,共用了 117 个样本点,拟合二阶响应面模型,得到的结论如表 38.3-13~表 38.3-16 所示。

表 38.3-13 响应  $y_1$ (MSE 均方误差)二阶模型的方差分析

Source	DF	平方和	均方
回归	36	58.34331	1.62065
残差	80	0.02334	2.917187E-04
未修正和	116	58.36665	
修正和	115	0.38416	

R 平方=1-残差 SS/修正 SS=0.93925

表 38.3-14 响应  $y_2$ (重量)二阶模型的方差分析

Source	DF	平方和	均方
回归	36	1354.63617	37.62878
残差	80	0.03278	4.097733E-04
未修正和	116	1354.66895	
修正和	115	0.38542	

R 平方=1-残差 SS/修正 SS=0.91494

表 38.3-17 几种模型的均方误差值比较表

系统响应	线性模型	交叉项模型	纯二次项模型	二阶模型
精度(MSE)	0.0431	0.0327	0.0348	0.0170
重量	0.0223	0.0198	0.0227	0.0201
基频	0.2015	0.1937	0.2026	0.1928
最大应力	2.4272	2.0867	2.2658	1.8635

表中粗体显示的数据为对每个响应面模型均方误差最小值。由此,采用的各系统响应的响应面模型确定为

表 38.3-15 响应  $y_3$ (基频)二阶模型的方差分析

Source	DF	平方和	均方
回归	36	175474.46962	4874.29082
残差	80	3.00331	0.03754
未修正和	116	175477.47293	
修正和	115	29.33688	

R 平方=1-残差 SS/修正 SS=0.89763

表 38.3-16 响应  $y_4$ (结构中最大应力)二阶模型的方差分析

Source	DF	平方和	均方
回归	36	285559.35297	7932.20425
残差	80	281.25751	3.51572
未修正和	116	285840.61049	
修正和	115	2180.23920	

R 平方=1-残差 SS/修正 SS=0.87100

由上面几个表格中的结果,我们发现四个二阶响应面模型也拟合得很好。为了从这些模型中取出最好的模型,这里比较了对每个响应拟合的线性模型、带交叉项的模型、带纯二次项的模型和二阶模型的均方误差值,见表 38.3-17。

$$\begin{aligned}
 y_1 = & 0.6259 - 0.0127x_1 - 0.0236x_2 - 0.0070x_3 - 0.0240x_4 - 0.0046x_5 + 0.0159x_6 + 0.0080x_7 - \\
 & 0.0007x_1x_2 + 0.0006x_1x_3 + 0.0010x_1x_4 + 0.0021x_1x_5 + 0.0011x_1x_6 - 0.0021x_1x_7 + \\
 & 0.0004x_2x_3 - 0.0007x_2x_4 + 0.0007x_2x_5 - 0.0152x_2x_6 + 0.0050x_2x_7 + \\
 & 0.0006x_3x_4 - 0.0016x_3x_5 + 0.0028x_3x_6 + 0.0010x_3x_7 - 0.0018x_4x_5 + \\
 & 0.0156x_4x_6 - 0.0092x_4x_7 + 0.0032x_5x_6 + 0.0016x_5x_7 + 0.0200x_6x_7 + \\
 & 0.0007x_1^2 + 0.0002x_2^2 + 0.0068x_3^2 + 0.0139x_4^2 - 0.0017x_5^2 + 0.0632x_6^2 + 0.0058x_7^2
 \end{aligned} \tag{38.3-75}$$

$$\begin{aligned}
 y_2 = & 3.4179 + 0.0162x_1 + 0.0239x_2 + 0.0266x_3 + 0.0150x_4 - 0.0273x_5 - 0.0125x_6 + 0.0159x_7 - \\
 & 0.0009x_1x_2 - 0.0009x_1x_3 + 0.0034x_1x_4 - 0.0009x_1x_5 - 0.0022x_1x_6 + 0.0039x_1x_7 + \\
 & 0.0005x_2x_3 - 0.0007x_2x_4 + 0.0018x_2x_5 - 0.0016x_2x_6 + 0.0026x_2x_7 + \\
 & 0.0023x_3x_4 - 0.0035x_3x_5 - 0.0013x_3x_6 - 0.0011x_3x_7 + \\
 & 0.0016x_4x_5 - 0.0000x_4x_6 + 0.0014x_4x_7 + \\
 & 0.0033x_5x_6 - 0.0033x_5x_7 + 0.0100x_6x_7
 \end{aligned} \tag{38.3-76}$$

$$\begin{aligned}
 y_3 = & -38.9410 + 0.2150x_1 + 0.1059x_2 - 0.0031x_3 + 0.2183x_4 - 0.0202x_5 - 0.3315x_6 - 0.1615x_7 - \\
 & 0.0120x_1x_2 - 0.0030x_1x_3 + 0.0152x_1x_4 - 0.0214x_1x_5 - 0.0205x_1x_6 - 0.0126x_1x_7 + \\
 & 0.0005x_2x_3 - 0.0129x_2x_4 + 0.0231x_2x_5 + 0.0158x_2x_6 + 0.0045x_2x_7 + \\
 & 0.0229x_3x_4 - 0.0048x_3x_5 + 0.0148x_3x_6 - 0.0019x_3x_7 + \\
 & 0.0102x_4x_5 + 0.0100x_4x_6 + 0.0933x_4x_7 + \\
 & 0.0034x_5x_6 - 0.0017x_5x_7 - 0.0242x_6x_7 + \\
 & 0.1069x_1^2 - 0.0994x_2^2 - 0.0785x_3^2 + 0.0496x_4^2 + 0.0820x_5^2 + 0.0163x_6^2 + 0.0901x_7^2
 \end{aligned} \tag{38.3-77}$$

$$\begin{aligned}
 y_4 = & 46.2637 - 1.9406x_1 + 0.2931x_2 + 0.1617x_3 - 0.4237x_4 + 0.2806x_5 + 0.7416x_6 - 3.1942x_7 + \\
 & 0.1405x_1x_2 - 0.1653x_1x_3 - 0.1554x_1x_4 - 0.1675x_1x_5 - 0.1070x_1x_6 - 1.3685x_1x_7 + \\
 & 0.2066x_2x_3 - 0.1132x_2x_4 - 0.2734x_2x_5 - 0.0984x_2x_6 + 0.1403x_2x_7 + \\
 & 0.1145x_3x_4 + 0.1580x_3x_5 - 0.2280x_3x_6 - 0.1464x_3x_7 + \\
 & 0.4033x_4x_5 + 0.1324x_4x_6 + 0.0809x_4x_7 + \\
 & 0.1218x_5x_6 - 0.0019x_5x_7 + 0.4558x_6x_7 + \\
 & 1.0775x_1^2 - 0.0804x_2^2 - 0.2803x_3^2 + 0.4674x_4^2 + 0.1690x_5^2 - 1.0117x_6^2 + 2.6364x_7^2
 \end{aligned} \tag{38.3-78}$$

各响应面模型的曲线由图 38.3-65~图 38.3-68 表示。

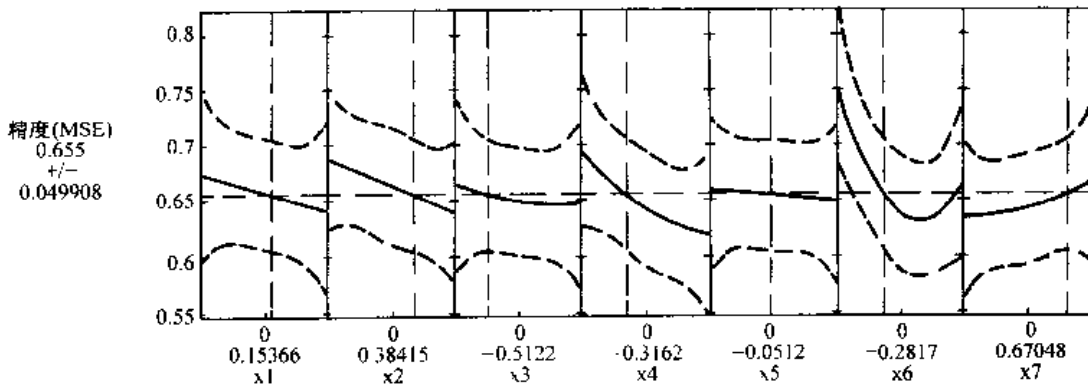


图 38.3-65 响应  $y_1$ (MSE 均方误差) 的响应面模型

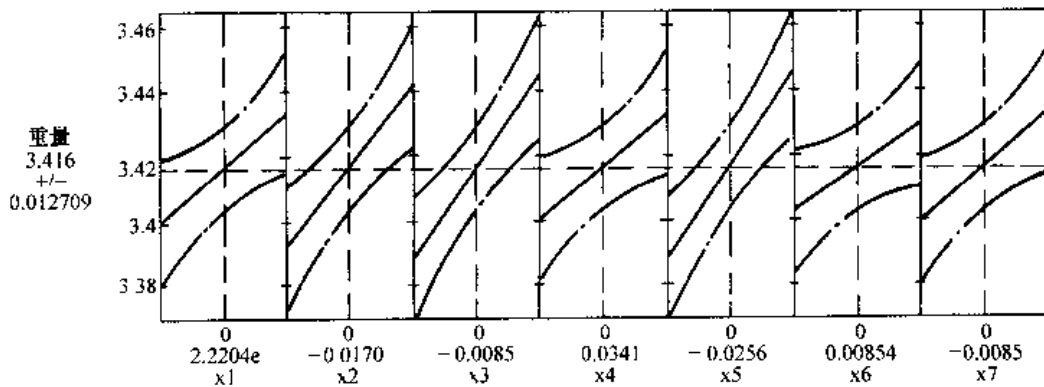


图 38.3-66 响应  $y_2$ (重量) 的响应面模型

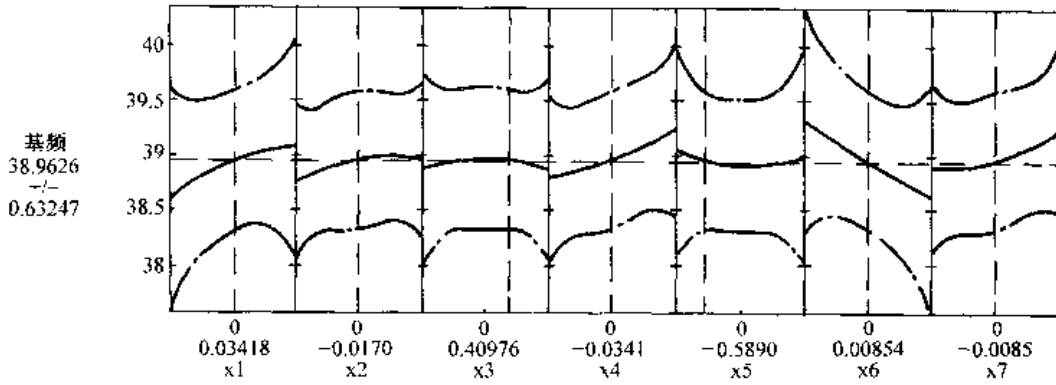


图 38.3-67 响应  $y_1$ (基频) 的响应面模型

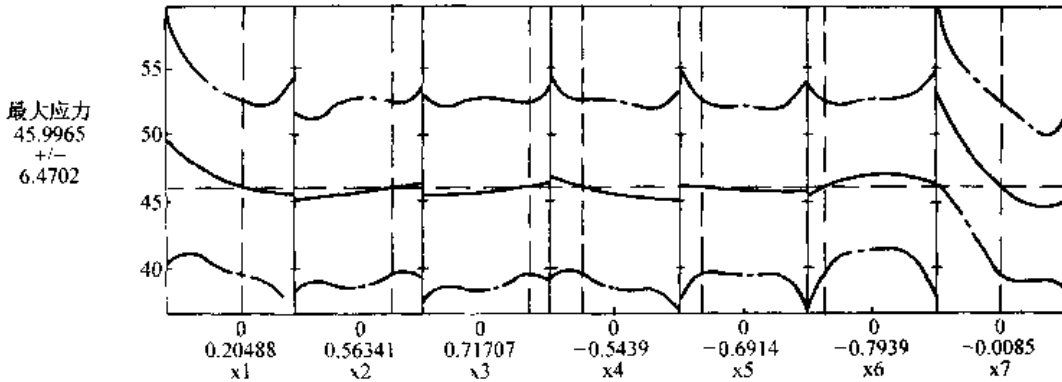


图 38.3-68 响应  $y_4$ (最大应力) 的响应面模型

第 4 步, 建立健壮性设计数学模型, 求出  $\mu$  及  $\sigma^2$  的表达式

$$f(1) = \bar{y}_1 \quad (38.3-79)$$

$$f(2) = y_2 \quad (38.3-80)$$

$$f(3) = \sum_{i=1}^7 \left( \frac{\partial y_1}{\partial x_i} \right)^2 \sigma_i^2 \quad (38.3-81)$$

$$f(4) = \sum_{i=1}^7 \left( \frac{\partial y_2}{\partial x_i} \right)^2 \sigma_i^2 \quad (38.3-82)$$

对原问题的约束条件也要做相应的变换, 得到

$$g(1) = 12/f_0 - 1 + \Delta g_1 \leq 0 \quad (38.3-83)$$

$$g(2) = \sigma_{max}/\sigma_0 - 1 + \Delta g_2 \leq 0 \quad (38.3-84)$$

$$g(3) = y_1/0.7 - 1 + \Delta g_3 \leq 0 \quad (38.3-85)$$

其中,  $\Delta g_1$ 、 $\Delta g_2$  和  $\Delta g_3$  表示系统的变异, 由下式近似得到

$$\Delta g_1 = \sum_{i=1}^7 \left| \frac{\partial g_1}{\partial x_i} \Delta x_i \right| \quad (38.3-86)$$

$$\Delta g_2 = \sum_{i=1}^7 \left| \frac{\partial g_2}{\partial x_i} \Delta x_i \right| \quad (38.3-87)$$

$$\Delta g_3 = \sum_{i=1}^7 \left| \frac{\partial g_3}{\partial x_i} \Delta x_i \right| \quad (38.3-88)$$

其中  $\Delta x_i$  ( $i=1, 2, \dots, 7$ ) 表示可控因素的变异范围, 假设各可控因素在其设计空间中是均匀分布的, 则取各正规化后的可控因素的变异范围为 0.1, 即 10% 的不确定范围。这样, 允许各可控因素的取值空间为  $x_i - \Delta x_i$  到  $x_i + \Delta x_i$  ( $i=1, 2, \dots, 7$ )。

第 5 步, 用 Pareto 遗传算法求解健壮性设计问题

根据上一步建立的健壮性设计模型, 用 Pareto 遗传算法来求解该问题, 数学模型为

$$\begin{aligned} \min \{ & f(1), f(2), f(3), f(4) \} & (38.3-89) \\ \text{s. t. } & g(1) \leq 0 \\ & g(2) \leq 0 \\ & g(3) \leq 0 \end{aligned}$$

设计变量  $x_1, x_2, \dots, x_7$ , 其中  $x_1 \sim x_5$  为离散变量,  $x_6$  和  $x_7$  为连续变量。用作者编制的 MOGA 软件包求解这一健壮性设计问题, 计算参数为

- 群体规模: 60
- 进化代数: 200
- Pareto 解集过滤器规模: 80
- 交叉概率: 0.5
- 变异概率取: 0.008

求解得到 Pareto 最优解集和对应的最优设计变量集合。由于 Pareto 最优解集中包含了四个目标函数的组合,用二维或三维曲面无法表示。考虑各目标的重要程度时选取决策者最满意点的方法,取四个目标函数的权值相同,即取  $w_1 = w_2 = w_3 = w_4 = 1/4$ ,得到最满意解的设计变量及目标函数值分别为

$$x_1 = 1, x_2 = 1, x_3 = -0.5, x_4 = 0.5, x_5 = -1, x_6 = 0.2621, x_7 = 0.1659,$$

$$f(1) = 0.5821, f(2) = 3.1374, f(3) = 1.48 \times 10^{-5}, f(4) = 3.58 \times 10^{-5}$$

得到的设计变量的实际值为

- (1):  $\phi 50 \times 3$ , (2):  $\phi 50 \times 3$ , (3):  $\phi 65 \times 3.5$ ,
- (4):  $\phi 40 \times 1$ , (5):  $\phi 50 \times 3$ , (6):  $\phi 40 \times 2.5$ ,
- (7):  $\phi 65 \times 3$ , (8):  $\phi 40 \times 2$ , (9):  $\phi 50 \times 2$ ,
- (10):  $\phi 50 \times 2$ ,

$$y_A = 60.516\text{cm} \quad y_B = 140.364\text{cm}$$

比较原设计:

- (1):  $\phi 50 \times 3$ , (2):  $\phi 50 \times 3$ , (3):  $\phi 65 \times 4$ ,
- (4):  $\phi 40 \times 2$ , (5):  $\phi 50 \times 3$ , (6):  $\phi 40 \times 2$ ,
- (7):  $\phi 65 \times 4$ , (8):  $\phi 40 \times 2$ , (9):  $\phi 50 \times 3$ ,
- (10):  $\phi 50 \times 3$ ,

$$y_{A0} = 71\text{cm} \quad y_{B0} = 159\text{cm}$$

其中  $x_1 \sim x_7$  的取值为正规化后的取值,从结论中可以看出,经过健壮性设计,表征质量指标的目标函数  $f(3)$  和  $f(4)$  的值都很小,说明该健壮性设计能在系统

达到原设计目标的情况下尽可能减小可控因素的变异对系统性能的影响。

用求得的最满意设计变量的值转化为标准值后,用有限元程序进行验算。验算得到的结果为

$f(1) = 0.5912, f(2) = 3.4575$ 。其误差为对  $f(1)$ : 1.54%,对  $f(2)$ : 0.58%。从误差来看,这里采用的转换模型的精度是相当高的。

与该天线原设计比较,原设计的结论为:  $f_0(1) = 0.6237, f_0(2) = 3.4847$ 。比较的结果为:经过健壮性设计后,天线的精度提高 6.36%,重量略有减轻。说明健壮性设计既能使系统的性能更优,又能提高产品的质量,同时,该设计得到的天线参数能使天线承受较大的内部和外部变异的影响而保持性能基本不变,使天线具有健壮性。

### 第6步,建立天线的多目标优化问题

天线的多目标优化模型为

$$\begin{aligned} \min \{ & f(1), f(2) \} & (38.3-90) \\ \text{s.t. } & g'(1) = 12/f_2 - 1 \leq 0 \\ & g'(2) = \sigma_{\max}/\sigma_0 - 1 \leq 0 \\ & g'(3) = f(1)/0.7 - 1 \leq 0 \end{aligned}$$

用 MOGA 软件包进行计算后得到该多目标优化问题的 Pareto 最优解集,见图 38.3-64 所示。为了从中选出最满意解,考虑到决策者对两个目标的权衡,这里选取了两个目标的加权系数变化时两个目标的取值及对应的设计变量的取值,见表 38.3-18 所示。

表 38.3-18 考虑目标加权时设计变量及目标函数值

$w_1$	$w_2$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$f(1)$	$f(2)$
1	0	1	1	0.5	1	1	0.1535	0.5809	0.5671	3.529
0.9	0.1	1	1	0.5	1	1	0.1535	0.5809	0.5671	3.529
0.8	0.2	1	1	0.5	1	1	0.1535	0.5809	0.5671	3.529
0.7	0.3	1	1	0.5	1	1	0.1535	0.5809	0.5671	3.529
0.6	0.4	1	1	0.5	1	-1	0.2598	1	0.5698	3.4857
0.5	0.5	1	1	0	0.8	-1	0.1640	0.1907	0.5774	3.1493
0.4	0.6	1	1	-1	0.2	-1	0.0589	-0.9991	0.6082	3.3786
0.3	0.7	1	0.8	1	0.2	-1	-0.4844	-0.9333	0.6785	3.3036
0.2	0.8	-1	-1	-1	-0.8	-1	-0.1649	-1	0.6946	3.2947
0.1	0.9	-1	-1	-1	-0.8	-1	-0.1649	-1	0.6946	3.2947
0	1	1	-1	-1	-0.8	-1	-0.1649	-1	0.6946	3.2947

决策者可根据表中的权系数的组合来选择最满意的解。与原设计  $f_0(1) = 0.6237, f_0(2) = 3.4847$  相比较,取  $w_1 = 0.6, w_2 = 0.4$  的设计结论  $f(1) = 0.5698, f(2) = 3.4857$  与之相比,在重量基本不变的情况下,精度提高 9.46%,其效益是很明显的。将此结论与健壮性设计的结果比较,发现目标函数变异的两个目标函数值分别为  $f(3) = 1.373 \times 10^{-5}, f(4) = 3.397 \times 10^{-5}$ ,大大高于健壮性设计的结果。这说明用健壮性设计方法能得到对系统参数的变异不敏感的设计结果,

使得系统在其可控因素或不可控因素有较大变异的情况下仍能保持稳定的性能。

### 3.5.6 总结

本节用基于健壮性的产品及过程开发方法对 10m 圆抛物面天线进行健壮性设计及多目标优化设计,用对天线性能影响最大的主梁的 10 根梁的截面特性及两个主要梁高参数作为设计变量,研究了各参数对系统的四个响应的影响,并从中筛选出 7 个关键参数作



为主要因素进一步研究。采用序列试验设计方法,先拟合低阶响应面模型,又用可旋转的中心复合设计得到拟合二阶模型的样本点,通过比较,确定了系统四个相应的二阶响应面模型。由此得到了系统的关键输入变量和输出响应之间的转换模型。根据该模型,确定了求解多目标优化设计问题及健壮性设计问题的数学模型,并用发展出的 Pareto 多目标遗传算法进行了求解。通过比较,发现两种设计方法虽然从问题的形式上有某些相似之处,但实际上处理的是两种完全不同的问题。多目标优化设计目的是在系统设计的目标在受到约束条件的限制下得到最优解,这种方法仅仅强调系统的尽可能满足目标,却不重视系统性能在最优的情况下是否稳定,可能会导致得到的最优性能处于很不稳定的状态,系统的内部条件和外部条件稍有变化,系统性能就下降很多,甚至有可能失效。

而健壮性设计问题则是以系统的质量为目标,其方法是使系统在满足目标原有的目标函数的前提下,尽可能使由于可控因素或不可控因素的变异造成系统的变异极小化,这样得到的解能保证系统在受到内在因素或外界因素干扰的情况下仍能保持良好的性能。健壮性设计得到的解可能比多目标优化得到的解从优化的角度要差一些,但它强调系统的质量,这一点正是产品设计的关键。得到健壮性设计的解以后,在后期的详细设计阶段对产品的参数在小范围调整时,不需要

或较少需要对系统的性能重新校验,节约了设计的时间和人力资源;同时,由于健壮性设计能得到对系统的内部因素和外界条件变化不敏感的健壮解,使得对产品加工过程中的要求和使用条件的要求都不是很高,而且产品的部件可以采用较低的配置而仍能获得较高的质量,大大减小了产品设计和制造的成本,能够带来很好的经济效益和社会效益。

#### 4 供应链库存策略的进化重组

本节主要从层次化的观点来考虑供应链中的库存,分别从定性和定量的角度来分析存在于供应链中的与整体制造策略有关的供应链整体解耦库存设置和供应链运作过程中各节点的运作库存策略。

##### 4.1 供应链运行策略的持续改进

经过层次模块化的 Petri 网建模,可以对供应链的总体运行效果做出评价,并可以指导敏捷供应链的结构重组过程。当供应链的网络结构初步形成以后,各个企业之间就形成一种相对稳定的战略伙伴关系,通过相互信任的商业伙伴间的合作来协调供应链,降低库存水平,实现整个供应链计划与执行过程的集成,如图 38.3-69 所示。过程的关键是强调贸易伙伴间建立共享的商业计划、销售预测和定单预测,通过前后节点的协调,实现敏捷供应链的合理运作。

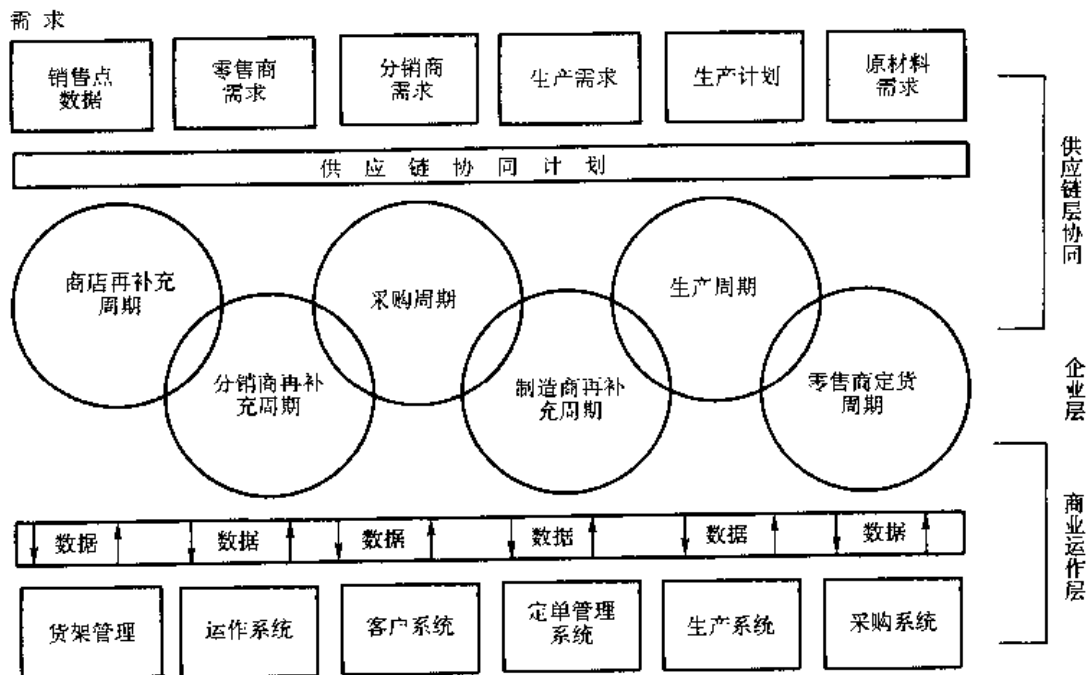


图 38.3-69 供应链运行过程中计划与执行过程的集成

在供应链的运行过程中,会发现许多需要改进的地方,类似于戴明的 PACD 循环(P(plan)——计划;D(do)——执行;C(check)——检查;A(action)——行动),供应链的运行过程也必须在运行过程中不断改进,图 38.3-70 所示为建立的供应链持续改进过程——ERMI 过程(E(evaluation)——评价;R(reengineering)——重组;M(monitor)——监控;I(improve)——改善)。

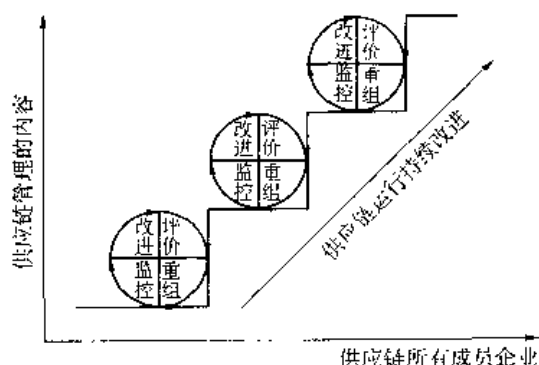


图 38.3-70 敏捷供应链运行中的 ERMI (evaluation-reengineering-monitor-improve) 过程

ERMI 过程可简述如下:

E——评价:根据供应链运行的总体要求和组织方针,建立供应链目标和过程;对现有供应链流程进行评价,识别运行中的浪费和问题,主要是要寻找产生浪费和问题的因果关系,进而探究其根本原因。

R——重组:理解了问题的因果关系和发现问题的根本原因后,对供应链的各项运行策略进行改进,包括各项功能、任务的分配以及过程的改进。

M——监控:对于供应链新的配置进行控制,建立监控机制对供应链的运作进行持续的评估,不断发现其改进的空间。

I——改善:采取措施,以持续改进供应链过程,这是 ERMI 的最终目的。

ERMI 中的四个过程不是运行一次就结束,而是周而复始地进行,一个循环完了,解决一些问题,未解决的问题进入下一个循环,这样阶梯式上升的。可以对供应链中的需求控制策略、库存控制策略、供货控制策略、物流控制策略等进行不断改进,本文将只对供应链

运行过程中的库存策略进行探讨。

## 4.2 供应链中的库存设置

库存是指企业所有资源的储备。库存控制策略是指用来控制库存水平、决定补充时间及定货批量大小的整套制度和控制手段。

若从供应链的整体运行效果来考虑,除了供应链各个节点企业内部的原材料、零部件和产成品库存外,相对于不同的产品供应链和不同运行策略,在供应链节点之间有时还应该设置一定的库存以协调整条链的运行,这里称之为战略解耦库存。

精益生产和敏捷制造是近几年来兴起的先进制造思想。将这两种思想合理地融入供应链管理中也会有效地提高供应链的效率,但究竟采取何种策略应由供应链的结构、产品或服务的性质和市场环境来决定。

我们粗略地把供应链的结构按照不同的市场需求分为按订单设计(ETO:engineer to order)、按订单制造(MTO:make to order)、按订单组装(ATO:assemble to order,或 configure to order)、按库存生产(MTS:make to stock)等几种形式。

几种供应链结构的主要区别在于链中按定单生产和按预测进行生产的分界点(我们称之为解耦点)的不同。为了使供应链在解耦点处很好地连接,减少由于预测需求和实际需求的差异所带来的供应链效率的降低,通常要设置一库存——解耦库存来减小由于需求的不确定性所带来的波动。图 38.3-71 为各种制造策略所对应的解耦库存的位置(图中“制造/装配商”代表了供应链中一个或多个商业伙伴成员)。可见解耦库存的下游是由顾客需求拉动的,而上游的供应链则是由预测驱动的。“推”式供应链与“拉”式供应链的主要区别如表 38.3-19 所示。

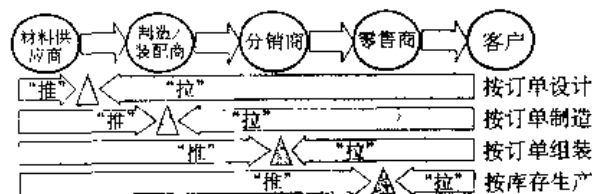


图 38.3-71 供应链中的解耦库存的位置

表 38.3-19 “推”式与“拉”式供应链的主要区别

	“推”式供应链	“拉”式供应链
预测依据	市场需求驱动预测	销售点(POS:point of sale)数据收集
生产方式	按库存生产	按订单生产
定货点	根据库存量(安全库存水平)确定定货点	根据客户需求自动、实时再补充
生产周期	较长的生产周期	较短的生产周期
市场情况	需求较为稳定,客户对产品多样性要求低	需求不稳定,客户对产品的多样性要求高

以色列物理学家 E. M. Goldratt 在他所著的《目标》一书中以小说的形式首次提出了约束理论(Theory of Constraints),并描述小说中的企业如何运用 TOC 使工厂转亏为盈。TOC 认为,任何一个系统都至少存在一个约束(有时也被称为“瓶颈”),要提高一个系统的产出,必须打破系统的约束。

TOC 的计划与控制是通过 DBR 系统实现的,即“鼓(drum)”、“缓冲器(buffer)”和“绳子(rope)”系统。“鼓”是指系统的约束;“缓冲器”是为了最大限度地利用瓶颈资源而设置的;非约束资源的产出应由约束资源的产出来控制,即“绳子”。整个系统按照“鼓的节拍”来运行。

供应链可以看成是一个环环相扣的链条,这个系统的强度取决于最弱的一环,我们要保证此约束资源能力的最大发挥,因为瓶颈资源损失 1h 相当于整个系统损失 1h,并且是无法补救的。

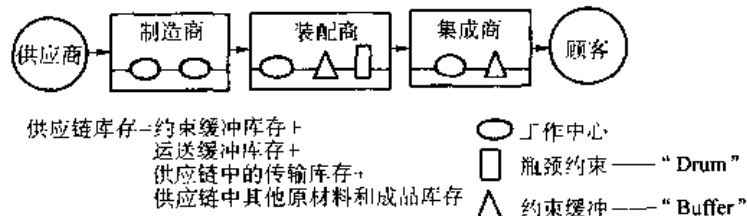


图 38.3-72 供应链中的缓冲库存

以上从战略的层次定性地探讨了供应链中的库存设置问题,在更低层次上的运作过程中,供应链中还会存在各种原材料、成品和在制品库存等,这些库存应在满足客户服务水平的基础上尽可能地减少,合理的库存控制策略将有助于供应链的协调运行和提高客户服务水平。以下将要从定量的角度探讨存在于供应链中的库存控制策略。

### 4.3 供应链运行过程中的库存控制策略

供应链中的企业持有存货的原因主要有:满足预测的需求、降低订货成本、减少缺货成本、使生产作业更为平稳和有弹性;但持有存货也有以下缺点:增加持有成本、难于对顾客需求的改变做出快速反应、占用企业大量资金等。在利弊共存的情况下,供应链中的存货控制的目的是以最小存货成本和满足客户需求为目标,使供应链中各节点不至于发生存货过多或是不足的情况。

在供应链竞争的年代,如何采取有效的库存控制策略以满足消费者持续变动并具有不确定性的需求,是供应链库存控制的首要任务。如何使供应链伙伴之间在利益共享、风险共担的协约下,以合作的方式来共同降低由于过量库存带来的成本浪费和由于缺货造成

为了保证约束资源的最大利用率,我们在约束位置设置一缓冲库存,确保约束资源不因材料短缺而闲置,进而保证整个系统的最大生产率。图 38.3-72 所示装配商中设置的库存即为约束缓冲。图 38.3-72 中还存在另外一种缓冲库存——运送缓冲,运送缓冲库存设置于系统的末端以确保准时运送。

为了确保供应链整体利益的最大化,供应链伙伴应合作识别系统的约束,将约束缓冲库存设置在该约束的前面,将运送缓冲库存设于整条供应链的末端。约束缓冲库存和运送缓冲库存水平可控制在一定范围之内,而其他即不存在约束缓冲又不存在运送缓冲的供应链伙伴则尽量将自身的库存排出。系统约束的生产率应尽量与市场的实际需求相接近,供应链其他伙伴则应尽量与系统约束的生产率同步。这种供应链伙伴间的协调和信任将使供应链产生最大效益。

的负面影响是有待研究的主要问题。

这里研究的目的是探讨供应链伙伴的合作策略,研究如何在有效的合作机制和策略下,有效地降低供应链总的库存持有成本、降低缺货率,取得供应链取得成本与顾客需求满意的平衡。本节将建立以成本和客户满意为主要目标的多目标模型,用遗传算法进行供应链成员最佳库存控制策略的搜寻,为供应链各成员的库存控制决策提供指导。

#### (1) 供应链中的库存类型及其作用

1) 库存的类型 根据物品需求的重复程度,可将库存分为单周期库存和多周期库存。单周期库存又称为一次性订货,即需求是一次性的,如报纸、月饼等;而多周期库存是在需求反复发生时,库存需要不断补充。这里主要研究多周期库存。

多周期需求库存又可分为独立需求库存与相关需求库存。独立需求的数量与出现的概率是随机的、不确定的和模糊的;相关需求的需求数量和需求时间与其他变量存在一定的相互关系,可以通过一定的数学关系推算得出。

2) 库存控制策略分类 库存控制的主要工作可分为三项:各种库存项目要有多少存货数量?在计划水平期间内应该订购或是生产多少数量的存货?要在什

么时候订购或生产?因此库存管理的关键是订购/生产的数量大小和订购/生产时间间隔问题。根据数量大小与时间间隔,库存控制系统可分为五类,如图 38.3-69 所示:

① 固定盘查周期-固定订货数量—— $(t, Q)$ 策略:如图 38.3-73a 所示,以订货周期和经济订货批量为标准的订货方式。每隔一定时期检查一次库存,并发出一次订货,订货量为固定值  $Q$ 。

② 固定盘查周期-变订货数量—— $(t, S)$ 策略:如图 38.3-73b 所示,以订货周期和最高库存量为标准。每隔一定时期检查一次库存,并发出一次订货,把现有库存补充到最大库存水平  $S$ ,若库存盘查时库存量为  $I$ ,则订货量为  $(S-I)$ 。

③ 连续盘查-固定订货数量—— $(R, Q)$ 策略:如图 38.3-73c 所示,以订货点(reorder point)和经济订货批量(economic lot size)为控制标准,即对库存进行连续盘查,当库存水平降至订货点水平  $R$  时,进行订货,订货量为固定值  $Q$ 。适用于需求量大、缺货费用较高、需求波动性大的情形。

④ 连续盘查-变订货数量—— $(R, S)$ 策略:如图 38.3-73d 所示,对库存进行连续盘查,当发现库存降低至订货点水平  $R$  时,进行订货,订货后使最大库存水平保持不变,即为  $S$ 。若发出订单时库存量为  $I$ ,则其订货量即为  $(S-I)$ 。与  $(R, Q)$ 策略不同的是,订货量按实际库存而定,订货量可变。

⑤  $(t, R, S)$ 策略:如图 38.3-73e 所示,  $(t, S)$ 和  $(R, S)$ 策略的综合,即设立一定的检查周期,若检查时发现库存量降至  $R$  时,即进行订货,订货量为设定的最大库存量  $S$  与现存量  $I$  之差  $(S-I)$ 。

#### (2) 支持供应链管理的库存控制策略

支持供应链管理的库存控制策略主要有供应商管理用户库存(VMI: vendor Managed Inventory)、联合库存管理系统(马士华, 2000)、多级库存模型(multi-echelon Inventory Model)。把供应链作为一个整体考虑时需采用集中控制,采用集中库存控制策略的一个重要方面是要获取供应链中其他企业的库存信息。

供应商管理用户库存(VMI: vendor managed inventory)(Holmström 1998; Achabal 2000): VMI 是一种在用户和供应商之间的合作性策略,以对双方来说都是最低的成本优化产品的可获性,在一个相互同意的目标框架下由供应商管理库存,以产生一种连续改进的环境。VMI 是供应链中的两个节点之间的交易模式,它总是能够改善供应链中的渠道成本和提高买方的利润,而对于卖方来说,短期内有时会有采购成本和利润的降低,通过长期的运行才会得到改善(Dong,

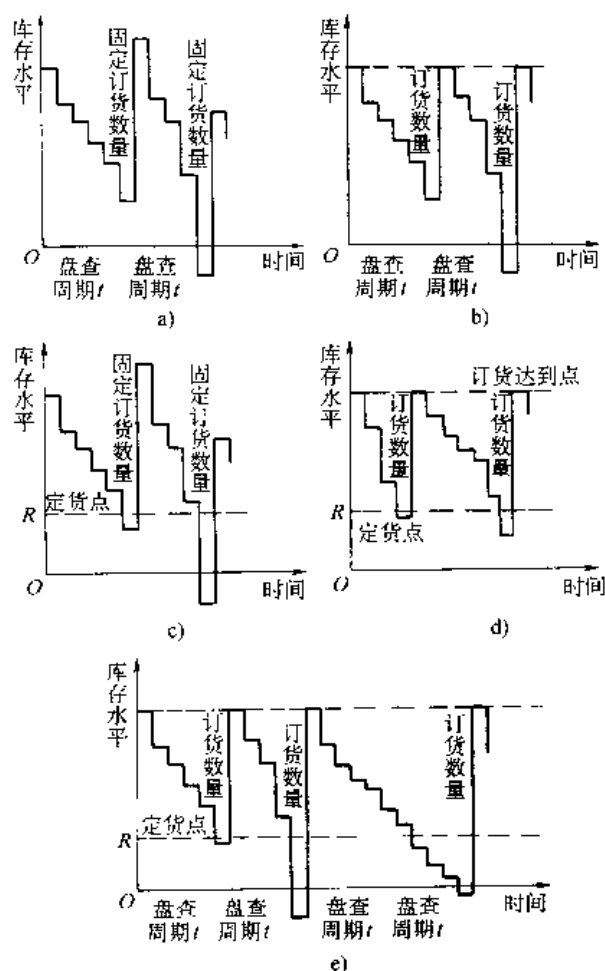


图 38.3.73 五种基本的库存控制策略  
a)  $(t, Q)$ 策略 b)  $(t, S)$ 策略 c)  $(R, Q)$ 策略  
d)  $(R, S)$ 策略 e)  $(t, R, S)$ 策略

2002)。

联合库存管理是提高供应链同步化程度的一种有效的方法,与供应商管理用户库存策略不同的是:联合库存管理强调双方同时参与,共同制定库存计划,保持供应链相邻节点之间的库存管理者对需求的预测保持一致,从而改善需求变异放大现象。任何相邻节点的需求的确定都是供需双方协调的结果,部分消除了由于供应链环节之间的不确定性和需求扭曲现象导致的供应链库存波动(马士华, 2000)。

Myers 定性地研究了在制造/销售两级供应链中,使用自动库存补充对提高顾客服务水平和降低成本的影响(Myers, 2000); Dejonckheere 研究了根据指数平滑得到的订单传递模式和库存控制反馈系统所产生的长鞭效应并进行仿真,指出“长鞭效应”这种供应链中的上游企业的订单波动总是大于实际的顾客需求的波动现象,可以通过合理的设计供应链和对供应链的重组,建立合理的物流控制过程来缓解(Dejonckheere,

2002)。

Larsena 研究了生产/分销系统的物流(Larsena, 1999);Ng 研究了由两个仓库和两个零售商组成的供应链,假定零售商面临的是独立的泊松分布需求、每个节点都使用 $(R,Q)$ 订货策略、外部供应能力无限的情况下的订购网络模型;Disney 建立需求随时间变化、卖方管理库存情况下的生产或分销计划模型——基于自动渠道、库存和订单的生产控制系统(APIOBPCS; automatic pipeline, inventory and order based production control system)(Disney, 2002);Ganeshan 研究了与多个供应商和多个零售商相连的中心仓库的生产/分销网络的接近最优的 $(R,Q)$ 库存控制策略模型(Ganeshan, 1999);Bhattacharjee 研究在供应链零售阶段,单个产品在生命周期的特定阶段的多阶段库存和价格模型(Bhattacharjee, 2000);Gavirneni 系统信息共享情况下的生产/分销系统的合作模型(Gavirneni, 2001);Boyaci 考虑出一个分销商和一个或多个零售商组成的供应链的协作问题。假定顾客需求随价格波动而相应变化情况下,定价和生产批量决策(Boyaci, 2002);Anderson 一个中心仓库和几个不同的零售商组成的供应链的库存控制问题(Anderson, 2000)。

供应商管理用户库存和联合库存管理策略都是对供应链的局部优化控制,而要进行供应链的全局性优化与控制,则必须采用多级库存控制与优化方法,以实现供应链资源的全局性优化。多级库存控制的方法有两种:集中式控制(centralized control)和分散式控制(decentralized control)。分散式库存控制是各个成员收集有效的部分相关信息独立制定库存控制决策,这种方式管理上相对简单,但容易产生次优的结果;集中式库存控制是集成供应链中的所有信息(如库存信息、需求信息等)来制定供应链的最佳化库存控制决策。集中式库存控制中,各库存点的控制参数是同时产生的,考虑了各节点之间的相互关系,通过协调的方法获取供应链库存的全局优化,因而对信息的交流和管理上的协调要求很高。

Petrovic 研究了不确定环境下供应链的模糊建模和仿真。以成本优化为目标,对供应链的运作进行动态仿真,来指导供应链中的库存决策(Petrovic, 1998)。徐贤浩等提出了一种供应链网络状结构模型中多级库存控制模型(徐贤浩, 1998),引入供应率和需求率两个参量,提出了供应链上各节点企业在保证生产、供应连续进行的条件下的最佳订货批量和最佳订货周期的确定方法,使得供应链总的库存费用最低。在模型的建立过程中,做了相应的假设,如假设上级供应商提供的零配

件(或货物)必须保证下一级供应商生产的连续进行,即不允许缺货;假设市场对核心产品的需求量在一段时间内是连续的、稳定的;假设供应商供应的货物(或零配件)的数量在一段时间内是连续的、稳定的等等。

以下将研究适应于敏捷供应链的库存协调模型,其中考虑供应链的成本以及对顾客的服务水平等评价指标。

#### 4.4 敏捷供应链多级库存策略重组模型

为了适应快速变化的市场环境对供应链敏捷性的要求,供应链各节点的库存策略需要根据市场需求和决策者对供应链的绩效需求而变更,建立图 38-3-74 所示的供应链库存策略设计和重组过程;首先定义供应链的网络结构和物料流、信息流等供应链过程,设定供应链运作的多个目标,如成本、顾客需求满意度等,建立供应链的优化模型并进行优化过程,得到基于供应链整体绩效考虑的各节点的理想库存策略,然后可根据决策者的意图对所设计的策略进行评价,如果满足要求则进行实施,否则进行下一轮的优化过程。

##### (1) 供应链总体运行策略模型

供应链中的运行策略包括了许多方面,如供应链成员间的利益分配、各节点的生产规划、产品促销、产品组合策略等。这里主要研究在产品种类相对稳定情况下的供应链各成员间的产品供应流。以供应链的总体成本和柔性作为评价指标。

1) 供应链总成本模型 供应链的总成本包括四部分,一是从供应商到工厂的原材料采购成本和运输成本;二是工厂生产的固定成本和变动成本;三是分销中心搬运和库存产品的变动成本和从工厂到分销中心的运输成本;四是从分销中心到顾客区的运输成本。供应链总成本模型:

$$Z = \left[ \sum_{r,j} (a_{r,j} + \lambda_{r,v}) A_{r,j} \right] + \left[ \sum_j f_{2j} q_{2j} + \sum_{ij} U_{2ij} X_{ij} \right] + \left[ \sum_k f_{3k} q_{3k} + \sum_{ikm} U_{3ik} D_{im} y_{km} + \sum_{ijk} c_{ijk} C_{ijk} \right] + \left[ \sum_{ikm} d_{ikm} D_{im} y_{km} \right] \quad (38.3-91)$$

2) 供应链柔性表达 供应链柔性在这里主要考虑为生产能力柔性和分销能力柔性,其中工厂的生产柔性用生产能力和生产能力利用之差描述,分销柔性用现实的分销量和顾客需求之差描述。供应链柔性模型:

$$W = \left[ \sum_j (q_{2j} \Phi_j - \sum_i \delta_{2ij} X_{ij}) \right] \omega_2 / \sum_j q_{2j} \Phi_j + \left[ \sum_k (q_{3k} \beta_k - \sum_{im} \delta_{3ik} D_{im} y_{km}) \right] \omega_3 / \sum_k q_{3k} \beta_k \quad (38.3-92)$$

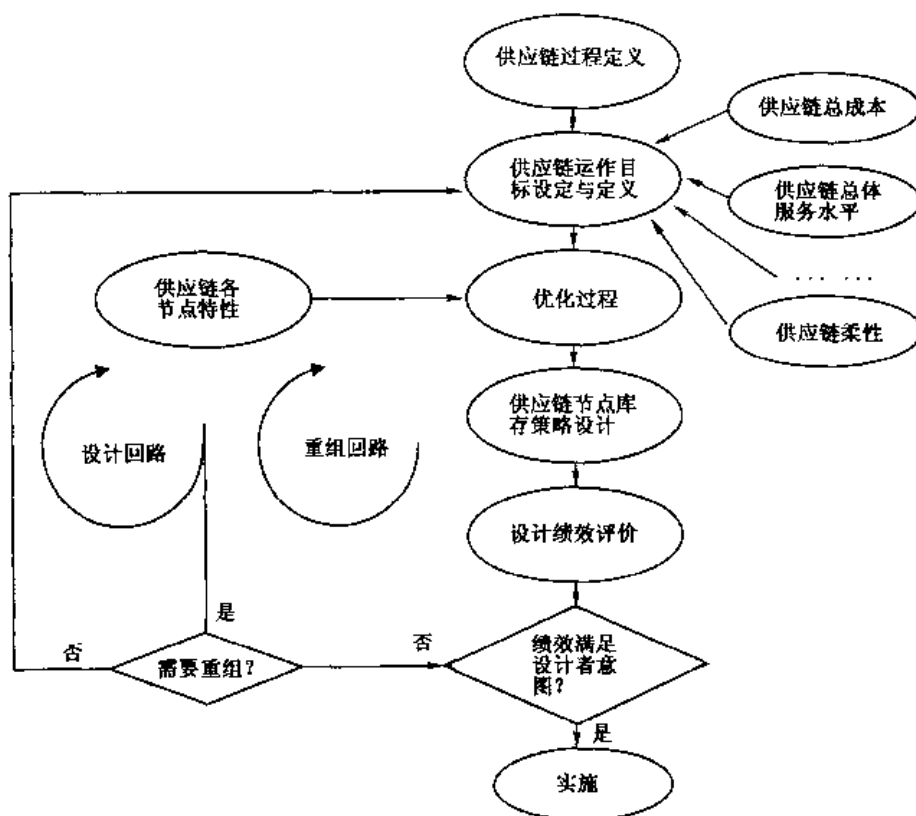


图 38.3-74 敏捷供应链库存策略设计和重组过程

供应链系统收到的约束条件如下：

$$\sum_j A_{rvj} \leq \Psi_{rv} \quad \forall r, v \quad (38.3-93)$$

$$\sum_i \tau_{rv} X_{ij} \leq \sum_v A_{rvj} \quad \forall r, j \quad (38.3-94)$$

$$\sum_j \delta_{2j} X_{ij} \leq \Phi_j q_{2j} \quad \forall j \quad (38.3-95)$$

$$\xi_{ij} q_{2j} \leq X_{ij} \leq \zeta_{ij} q_{2j} \quad \forall i, j \quad (38.3-96)$$

$$a_k q_{3k} \leq \sum_m \delta_{3mk} D_m y_{km} \leq \beta_k q_{3k} \quad \forall k \quad (38.3-97)$$

$$\sum_k y_{km} = 1 \quad \forall m \quad (38.3-98)$$

$$X_{ij} = \sum_k C_{ijk} \quad \forall i, j \quad (38.3-99)$$

$$\sum_{jk} C_{ijk} = \sum_m D_m \quad \forall i \quad (38.3-100)$$

$$\sum_j C_{ijk} = \sum_m y_{km} D_m \quad \forall i, k \quad (38.3-101)$$

$$X_{ij}, C_{ijk}, A_{rvj} \geq 0 \quad \forall i, v, j, k \quad (38.3-102)$$

$$q_{2j}, q_{3k}, y_{km} = 0 \text{ 或 } 1 \quad \forall j, k, m \quad (38.3-103)$$

其中：

式(38.3-93)是原材料供应限制；

式(38.3-94)是原材料运输限制；

式(38.3-95)是生产约束；

式(38.3-96)是工厂的生产数量控制；

式(38.3-97)保证分销中心的分销数量在最大分销规模与最小分销规模之间；

式(38.3-98)保证每个顾客区都分布有一个分销中心；

式(38.3-99)保证从工厂运输的产品数量与工厂的生产数量相等；

式(38.3-100)保证所有的需求都得到满足；

式(38.3-101)保证每个顾客区的需求得到满足。

在实际中，常常把供应链柔性作为约束条件来处理，即决策者在[0,1]范围选择适当的柔性期望值  $\epsilon$ ，然后令

$$W \geq \epsilon \quad (38.3-104)$$

这样，供应链战略优化问题可以表述为，在约束条件式(38.3-93)~式(38.3-104)条件下求取目标函数(1)的最小值，即在一定的产品需求组合下，求解一定时期内工厂向各个分销中心运送的产品总量  $C_{ijk}$ 、供应商向工厂运送的产品总量  $A_{rvj}$ 、工厂生产各产品的产量  $X_{ij}$  以及分销中心的设置  $q_{2j}$  和各分销中心对相应顾客区的服务  $y_{km}$  等。一般来说，随着供应链柔性要求

的增高,成本呈一定的上升趋势,决策者便需要在此两者之间作出权衡。

(2) 供应链库存策略的优化

通过供应链的总体运行策略的求解可确定一定时期内供应链成员的生产数量以及各成员间的总体物流情况,下面以一个由多个仓库组成的供应链系统中的多级库存控制策略为例,研究供应链的库存策略选择和优化。

考虑由一个主机企业(生产商)、一个分销商和  $N$  个零售商仓库组成的分布式库存系统,假设各个仓库间可以进行信息的交流并可以相互补充货物,如图 38.3-75 所示。面对不同的顾客需求,主机企业选择相应的库存点进行供货,根据各个仓库的库存情况,当零售商的总库存量降至总订货点以下时,在订货点以下的零售商按规定的订货数量提出订货,否则,进行相互补充。各个零售商的合作机制为:当某零售商面临缺货,若其他零售商有多余的存货时,则由该零售商供应多余的存货给缺货的零售商,在此机制下,可以减少缺货零售商的缺货成本和拥有多余存货的零售商的存货成本,同时也减少了整个供应链为了应付零售商的紧急订单所付出的成本。参照刘利民建立的模型(刘利民,2002),对某种商品,对各个零售商仓库作如下假定(表 38.3-20):

表 38.3-20 供应链多目标模型符号说明

变量	定义
$I$	产品类型编号, $i=1,2,\dots,I$
$V$	供应商(供应链第 1 层)编号, $v=1,2,\dots,V$
$J$	制造商(供应链第 2 层)编号, $j=1,2,\dots,J$
$K$	分销中心(供应链第 3 层)编号, $k=1,2,\dots,K$
$M$	零售商(供应链第 4 层)编号, $m=1,2,\dots,M$
$R$	原材料类型编号, $r=1,2,\dots,R$
$a_{r,j}$	原材料 $r$ 从供应商 $v$ 到制造商 $j$ 的单件运输费用
$c_{i,k}$	产品 $i$ 从制造商 $j$ 到分销中心 $k$ 的单件运输费用
$d_{i,k,m}$	产品 $i$ 从分销中心 $k$ 到零售商 $m$ 的单件运输费用
$\lambda_{rv}$	原材料 $r$ 对供应商 $v$ 的成本
$U_{2,j}$	产品 $i$ 在制造商 $j$ 处的单件制造成本
$U_{3,k}$	产品 $i$ 在分销中心 $k$ 处的单件通过成本(库存和搬运成本)
$F_{2,j}$	某一时段工厂 $j$ 的固定成本
$F_{3,k}$	某一时段分销中心 $k$ 的固定成本
$\Psi_{rv}$	供应商 $v$ 生产原材料 $r$ 的能力
$\delta_{2,j}$	生产一件产品 $i$ 在工厂 $j$ 所需生产的标准件数
$\delta_{3,k}$	每一件产品 $i$ 在分销中心 $k$ 所需的标准件数
$\Phi_j$	每个工厂的标准产品生产能力
$\tau_{ri}$	单位产品 $i$ 对每种原材料 $r$ 的利用率
$\xi_{i,j}$	工厂 $j$ 产品 $i$ 的最小生产规模
$\zeta_{i,j}$	工厂 $j$ 产品 $i$ 的最大生产规模
$\alpha_k$	分销中心 $k$ 的最小分销量
$\beta_k$	分销中心 $k$ 的最大分销量
$X_{i,j}$	某一时段工厂 $j$ 生产的产品 $I$ 的数量(件)
$C_{i,k}$	某一时段分销中心 $k$ 从工厂 $j$ 订购的产品 $i$ 的数量(件)

(续)

变量	定义
$A_{r,j}$	某一时段制造商 $j$ 从原材料供应商 $v$ 订购的原材料 $r$ 的数量(件)
$D_{i,m}$	为顾客区 $m$ 对产品 $i$ 的平均需求
$Z$	总成本
$W$	供应链柔性
$\eta$	顾客需求满意水平绩效, $[0,1]$
$\gamma$	运送柔性绩效水平, $[0,1]$
$\varepsilon$	柔性期望值, $[0,1]$
$\omega_2, \omega_3$	能力利用的权重, $[0,1]$
$q_{2j}$	为工厂设置, 取值为 1 或 0
$q_{3j}$	分销中心设置, 取值为 1 或 0
$y_{km}$	分销中心 $k$ 对顾客区 $m$ 的服务, 取值 1 或 0

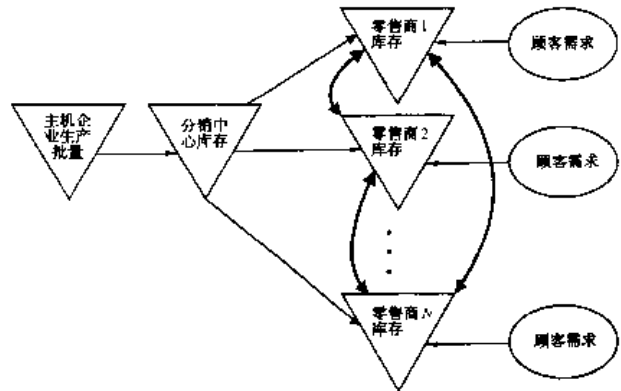


图 38.3-75 研究的供应链库存系统的物流结构

由主机企业进行统一库存策略规划,各个库存节点都采用连续盘查-固定订货量库存控制策略—— $(R,Q)$ 订货策略,订货允许所占的最大总金额  $CC$ ;

在  $t$  时段,每个零售商仓库的需求分别为  $D_{i,k}$ ,服从  $N(\mu_k, \sigma_k^2)$  的正态分布,  $k=1, \dots, N$ ;

每个仓库的最大库容分别为  $VV_k, k=1, \dots, N$ ,库存允许的最大库容  $CV_k$ ;

分销商到各仓库的提前期分别为  $L_k, k=1, 2, \dots, N$ ;

在  $t$  时段,制造商的供货能力为  $GQ_i$ ;

顾客需求满足率为  $P_r$ ;

假定主机企业到分销中心的提前期为  $L_m$ ;在  $t$  时段分销商的总需求为  $Q_t$ 。

优化目标:在资源和能力有限的情况下,确定各仓库的库存控制策略(安全库存、订货点、订货量),取得顾客满意度与库存总成本的权衡。

从顾客关系角度讲,过低的顾客需求满足率会导致顾客忠诚度(customer loyalty)的降低,这在市场竞争激烈的今天是尤为可怕的,因而以可接受的概率来保证顾客需求满足是非常重要的。首先,这里引入顾客需求满足率的表达,将供应链中各零售商库存的顾客需求满足率定义为现有库存能够满足订单的概率:

$$P_r = P[D_{i,k} \leq I_{i,k}] \quad (38.3-105)$$

考虑的供应链库存成本主要包括以下几项费用:

订货成本(ordering cost):与订购和进货有关的成本,主要指发出订单、验收货物和将货物移至仓库存储的相关的成本。订货成本一般与采购数量的关系不大,而与订货次数有关,因而这里假定订货成本的金额是固定的。

购货成本(purchasing cost):物料的采购成本与采购的数量有关,当采购数量达到一定范围时,物料的单价可能随采购量的增加而有所折扣,但也有涨价的可能。这里假定物料的单价购货成本是一定的。

储存成本(carrying cost):物料的储存成本包括资金积压损失、保险费、物料维护费等。

缺货成本(shortage cost):存货数量不能满足需求时所产生的损失。缺货成本通常可分为“可计算的有形缺货成本”和“不可计算的无形缺货成本”。前者包括停工待料的损失成本、延期交货的惩罚成本与销售损失的机会成本;后者包括信誉损失与顾客丧失产生的成本。

据此建立的供应链库存控制成本模型(表 38.3-21)

$$TC = TC_m + \delta \left( \sum_{k=1}^N R_k - \sum_{k=1}^N I_k \right) \times TC_1 + \delta \left( \sum_{k=1}^N I_k - \sum_{k=1}^N R_k \right) \times \left[ 1 - \prod_{k=1}^N \delta(I_k - R_k) \right] \times TC_2 \quad (38.3-106)$$

则建立供应链库存策略的多目标模型

$$\begin{aligned} \min TC \\ \max P_r \\ \text{s. t. } I_{k(t+1)} V_0 \leq VV_k \quad k=1,2,\dots,N \end{aligned} \quad (38.3-107)$$

$$\sum_{k=1}^N Q_{kt} \leq GQ_t \quad (38.3-109)$$

$$c1_{kt} \sum_{k=1}^N Q_{kt} \leq CC \quad (38.3-110)$$

$$V_c \sum_{k=1}^N I_{k(t+1)} \leq CV \quad (38.3-111)$$

其中:

$$TC_m = Q_t \times c0_t + [Q_t \times c1_t \times Q_t - \delta(Q_{kt}) \times c1_{kt} \times Q_{kt}] + c2_t \times (I_t + Q_t - D_t) \quad (38.3-112)$$

$$D_t = \sum_{k=1}^N D_{kt} \quad (38.3-113)$$

$$TC_1 = E \left( \sum_{k=1}^N CW_k \right) \quad (38.3-114)$$

$$TC_2 = E \left( \sum_{k=1}^N CWT_k \right) \quad (38.3-115)$$

$$I_{k(t+1)} = I_{kt} + Q_{kt} - D_{kt} \quad k=1,2,\dots,N \quad (38.3-116)$$

$$\begin{aligned} CW_k = & \delta(Q_{kt}) \times c0_{kt} + \delta(Q_{kt}) \times \\ & c1_{kt} \times Q_{kt} + c2_{kt} \times (I_{kt} + Q_{kt} - D_{kt})^+ + c3_{kt} \times \\ & (D_{kt} - I_{kt} - Q_{kt})^- \\ = & \delta(Q_{kt}) \times c0_{kt} + \delta(Q_{kt}) \times c1_{kt} \times Q_{kt} + \\ & c2_{kt} \times \sum_{D_{kt} \leq I_{kt} + Q_{kt}} ((I_{kt} + Q_{kt} - D_{kt}) \times p_k(D_{kt})) + \\ & c3_{kt} \times \sum_{D_{kt} > I_{kt} + Q_{kt}} ((D_{kt} - I_{kt} - Q_{kt}) \times p_k(D_{kt})) \\ k= & 1,2,\dots,N \end{aligned} \quad (38.3-117)$$

$$\begin{aligned} CWT_k = & \delta_2(Q_{kt}) \times c0_{kt} + c2_{kt} \times (I_{kt} + Q_{kt} - D_{kt})^+ + \\ & c3_{kt} \times (D_{kt} - I_{kt} - Q_{kt})^- + \\ & \sum_{j \neq k} (\delta_2(Q_{kj}) \times c4_{kj} \times Q_{kj} \times ds_{kj}) \\ = & \delta_2(Q_{kt}) \times c0_{kt} + c2_{kt} \times \\ & \sum_{D_{kt} \leq I_{kt} + Q_{kt}} ((I_{kt} - Q_{kt} - D_{kt}) \times p_k(D_{kt})) + \\ & c3_{kt} \times \sum_{D_{kt} > I_{kt} + Q_{kt}} ((D_{kt} - I_{kt} - Q_{kt}) \times p_k(D_{kt})) + \\ & \sum (\delta_2(Q_{kj}) \times c4_{kj} \times Q_{kj} \times ds_{kj}) \quad k=1,2,\dots,N \end{aligned} \quad (38.3-118)$$

式(38.3-118)中,  $Q_{kj} = \sum_{j=k}^N Q_{kj}$   $k, j=1,2,\dots,N$  (38.3-119)

另外,  $\delta(Q) = \begin{cases} 1 & Q > 0 \\ 0 & Q \leq 0 \end{cases}$  (38.3-120)

$\delta_2(Q) = \begin{cases} 1 & Q \neq 0 \\ 0 & Q = 0 \end{cases}$  (38.3-121)

零售商的安全库存量可表达为

$$ss_k = k_r \sigma_k \quad (38.3-122)$$

另外,订货点应包括订购提前期需求量加上期望服务水平下的安全库存量,两者关系可表达为

$$R_k = ss_k + L_k \times \bar{\mu}_k \quad (38.3-123)$$

$$1 - P_r = P_{\mu \geq k}(k_r) = 1 - \Phi(k_r) \quad (38.3-124)$$

### (3) 遗传算法求解过程

考虑一个由一个分销中心、三个零售商组成的供应链系统,按照上而的模型,设各个零售商仓库的参数为:提前期内的需求分布分别为  $N(150,150)$ 、 $N(170,170)$ 和  $N(220,220)$ ;初始库存量分别为 50、70 和 90;最大库存容量都为  $600m^2$ ,每件商品占用的库存  $1m^2$ ;交易费都为 120 元/次;购货费为 40 元/件;存储费为 0.5 元/件;提前期为 3 天;缺货损失为 200 元/件。分销商的各项参数为:初始库存量为 200;最大库存容量  $3000m^2$ ;交易费为 400 元;存储费 0.2 元/件;提前期为 3 天。考虑时段为 30 天。



供应链库存优化问题属于 NP-hard 问题,用传统的数学方法求解,过程复杂而且效率不高,用启发式算法求解效率高但解的质量难于保证。这里采用兼具求解效率与效果的多目标遗传算法。

用 EPEA 进行订货量的求解,主要参数确定通过以下方面来考虑:

**编码与解码:**遗传算法操作的对象为能表示问题解的字串,在进行演算算法之前要对问题进行编码,将问题的决策变量转化为固定长度的字串,字串中的符号相当于生物学上的基因,记录着生物个体的遗传特性,将这些符号排列成字串就形成了染色体。对本问题采用实数编码,每一仓库的订货量  $Q_k$  代表一个基因,根据约束条件可以进一步确定问题的可行域;

**初始化群体:**群体规模影响遗传优化的最终结果以及遗传算法的执行效率。当群体规模  $n$  太小时,遗传算法的优化性能一般不好,采用较大的群体规模可以减少遗传算法陷入局部最优解的机会,但大的群体规模意味着计算复杂度高,这里取  $n=20$ 。

**适应度函数确定:**适应度函数是用以评估原始问题的目标函数,通过适应度函数计算染色体的适应度值,并以此来提供判断染色体优劣的依据。适应度值较佳者,被选取以产生下一代的机率较高,而适应度较差的,则容易遭到淘汰,以此进行优生演化。对供应链库存控制问题,适应度是对所选策略有效性的测量,这里直接用成本函数表达。

**复制:**模拟自然界“适者生存,不适者淘汰”的法则,从染色体群中,利用适应度值的评估结果,将适应度程度较好者挑选并复制至交叉盘中,以准备演化产生新的染色体子代。

**交叉:**交叉率  $P_c$  控制着交叉操作被使用的频度。较大的交叉率可增强遗传算法开辟新的搜索区域的能力,但原本适应度值好的染色体却可能因此遭到破坏,而演化成适应度值较差的染色体,而失去优生演化的效果;若交叉率太低,遗传算法搜索可能陷入迟钝状态,这里取  $P_c=0.5$ 。

表 38.3-21 供应链库存策略优化模型符号说明

符号	定义
$I_{kt}$	零售商仓库 $k$ 在 $t$ 时段的初始库存, $k=1,2,\dots,N$
$I_{jt}$	分销商仓库在 $t$ 时段的初始库存
$p_k(j)$	零售商仓库 $k$ 在 $t$ 时段需求为 $j$ 的概率分布
$Q_{kt}$	零售商仓库 $k$ 在 $t$ 时段向分销商的订货量
$Q_{jt}$	分销商仓库在 $t$ 时段向主机企业的订货量
$R_k$	零售商仓库 $k$ 在 $t$ 时段的订货点
$R$	分销商仓库在 $t$ 时段的订货点
$ssa_k$	零售商仓库 $k$ 在 $t$ 时段的安全库存
$ss$	分销商仓库在 $t$ 时段的安全库存
$c0_{kt}$	零售商仓库 $k$ 在 $t$ 时段每次购货的订货成本
$c0_{jt}$	分销商仓库在 $t$ 时段每次购货的订货成本

(续)

符号	定义
$c1_{kt}$	零售商仓库 $k$ 在 $t$ 时段单位商品的采购成本
$c1_{jt}$	分销商仓库在 $t$ 时段单位商品的采购成本
$c2_{kt}$	零售商仓库 $k$ 在 $t$ 时段单位商品的储存成本
$c2_{jt}$	分销商仓库在 $t$ 时段单位商品的储存成本
$c3_{kt}$	零售商仓库 $k$ 在 $t$ 时段单位商品的缺货损失成本
$c3_{jt}$	分销商仓库在 $t$ 时段单位商品的缺货损失成本
$c4_{kt}$	零售商仓库 $k$ 在 $t$ 时段的单位运输费
$ds_{kj}$	零售商仓库 $k$ 到仓库 $j$ 的距离
$(W)_k$	零售商仓库 $k$ 的订货总费用
$(WT)_k$	零售商仓库 $k$ 的调拨总费用
$Q_{kj,t}$	零售商仓库 $k$ 在 $t$ 时段向仓库 $j$ 的调拨量(调入为正,调出为负)
$\mu_k$	零售仓库 $k$ 在某时段的需求均值
$\sigma_k^2$	零售仓库 $k$ 在某时段的需求方差
$\bar{\mu}_k$	零售仓库 $k$ 在单位时间的需求均值
$k_r$	零售商库存安全系数
$P_r$	顾客需求满足率
$P_{\mu \geq k}(k)$	标准正态分布中变量大于等于 $k$ 的概率
$\Phi(k)$	标准正态分布中变量为 $k$ 时的累积分布函数

**变异:**变异在遗传算法中属于辅助性的搜索操作,它的主要目的是维持群体的多样性。一般来说,低频度的变异可将染色体中导入其他的基因结构,以进行其他可行解区域的搜寻,但过高的突然变异率将使遗传算法趋于纯粹的随机搜索的求解方法,这里取  $P_m=0.05$ 。

在此条件约束下,采用前面建立的 EPEA 软件过程进行搜索,在顾客需求满意率为 96% 时的收敛曲线如图 38.3-76 所示。

经过进化得到的相应顾客需求满足率情况下的最优方案如表 38.3-22 所示。

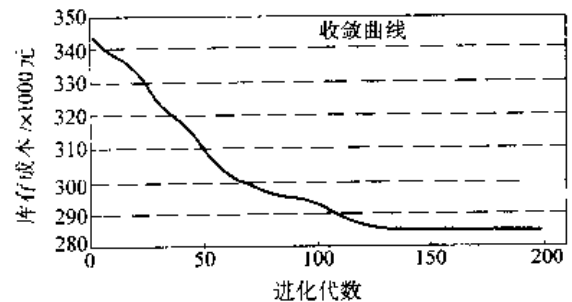


图 38.3-76 进化过程的收敛曲线

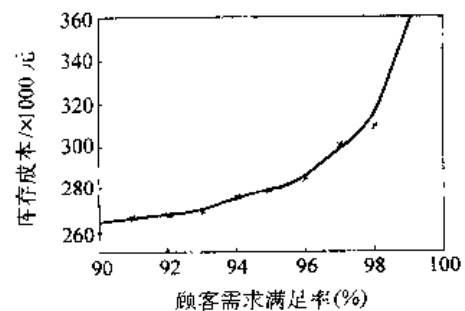


图 38.3-77 库存成本与顾客满足率关系曲线

表 38.3-22 一定顾客满足率下的最优方案

仓库	订货点	安全库存量	订货量
零售仓库 1	165	15	332
零售仓库 2	187	17	349
零售仓库 3	240	20	376
分销商仓库	570	30	1025
顾客需求满足率 90% 总成本:261 430 元			
零售仓库 1	171	21	342
零售仓库 2	193	23	362
零售仓库 3	246	26	395
分销商仓库	580	40	1248
顾客需求满足率 96% 总成本:284 740 元			
零售仓库 1	175	25	354
零售仓库 2	197	27	387
零售仓库 3	251	31	435

(续)

仓库	订货点	安全库存量	订货量
顾客需求满足率 96% 总成本:284 740 元			
分销商仓库	589	49	1452
顾客需求满足率 98% 总成本:308 940 元			

供应链的库存成本与顾客需求满足率的关系曲线如图 38.3-77 所示。

此时决策者可以在顾客需求满足率和总成本之间作出权衡,选择相应的库存补充策略。在供应链的运行过程中,随着外部需求的不断改变和企业特征如提前期的改变等,可以借此模型来对供应链中的库存策略进行改进和重组,以在一定的顾客满足率情况下使库存总成本最小。

## 第4章 自组织设计技术与方法

元胞自动机模型能十分方便地复制出复杂的现象或动态演化过程中的吸引子、自组织和混沌现象。一般来说,复杂系统由许多基本单元组成,当这些子系统或基元相互作用时,主要是邻近基元之间的相互作用,一个基元的状态演化受周围少数几个基元状态的影响。在相应的空间尺度上,基元间的相互作用往往是比较简单的确定性过程。用元胞自动机来模拟一个复杂系统时,时间被分成一系列离散的瞬间,空间被分成一种规则的格子,每个格子在简单情况下可取0或1状态,复杂一些的情况可以取多值。在每一个时间间隔,网格中的格点按照一定的规则同步地更新它的状态,这个规则由所模拟的实际系统的真实物理机制来确定。格点状态的更新由其自身和四周邻近格点在前一时刻的状态共同决定。不同的格子形状、不同的状态集和不同的操作规则将构成不同的元胞自动机。在一维模型中,是把直线分成相等的许多等分,分别代表元胞或基元;二维模型是把平面分成许多正方形或六边形网格;三维是把空间划分出许多立体网格。

以元胞自动机为基础的模型提供了完全不同的另外一种方法。在这个方法中,时间空间变量以至描述系统状态的变量都是分立的,它们所展示的自组织过程的复杂行为完全可以和微分方程或迭代映射所提供的相媲美。另一方面,由于元胞自动机所固有的特点,可用于描述更为复杂的现象。

### 1 自组织技术基础

#### 1.1 “生命的游戏”

1943年,数学家Pitts和神经心理学家McCulloch设想了一个由常规的神经元组成的系统,每个神经元相当于一个逻辑元胞,可以选择不同的逻辑值,而这些逻辑元胞可以在下一个时间一起将它们的逻辑值传递到别的逻辑元胞,那么,它们综合作用的结果就将是一个时间系统的结果,这种可以自动演化的机器就是所谓的图灵机(Turing)或有限状态机。元胞自动机最早是由冯·诺伊曼(Von Neumann)和乌拉姆(Ulam)于20世纪60年代提出的。冯·诺伊曼提出构造一个不确定的生命模型系统的设想,这个系统可以智能地自我演化。后来,冯·诺伊曼将这个模型发展为一个网格状的自动机网络,每个网格为一个元胞自动机,元胞状态有生和死,相当于人体组织的存活和消亡。但是,

早期的元胞自动机思想,许多问题处于未知状态。当时的名字叫元胞空间(cellular spaces),用于模拟生物学中的自复制。后来被用于研究许多其他现象,随之又出现了各种各样的名称,例如元胞结构(cellular structures)、镶嵌自动机(tessellation automata)以及元胞自动机(cellular automata)等等。

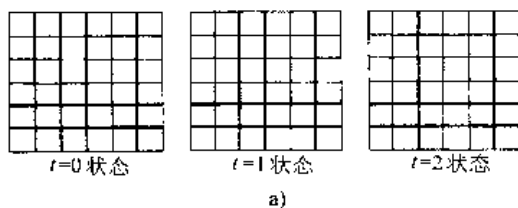
1970年,剑桥大学数学家Conway发明了一种叫做“生命的游戏”的游戏,这个游戏完全体现了动态元胞自动机的特征。游戏是这样的:如图38.4-1所示,游戏是在一张类似棋盘的平面网格上进行的,这些网格上可以放上棋子或不放,由游戏规则决定。棋子的意思代表有一个生命存活,而没有棋子的地方代表无生命存活。开始时,在网格上随机地摆上一些棋子,然后,按照一定的规则来确定每个有棋子位置上棋子的存在,以及没有棋子位置上是否加入新的棋子。规则是:每个棋子有可能有8个邻居,而有4个是直接相邻的,如果目前的棋子有2个或3个邻居,那么它将在下一次被考虑前保留;如果它有4个或4个以上邻居时,它将被认为是因为人口过多将会死亡而被取走;如果只有1个或没有邻居,那么将会因为过于寂寞而死,所以也将被取走。同时,如果当前位置上没有棋子,而此位置周围正好有3个邻居时,那么这个位置将被放上一个新的棋子。整个游戏就按这个规则进行下去,一代,一代……。

图38.4-1为在游戏过程中任取的三个连续状态。特别需要说明的是,处于边界的元胞,在判断时认为网格的外围存在一圈网格,但网格默认为是空白的(至于边界的处理方法将在后续节中讨论)。

从这个游戏中我们可以发现,元胞自动机系统中存在两个操作对象:元胞的状态和系统演化规则。其中元胞状态一般可以取二值变量,如0和1,其实在复杂的元胞自动机系统中,元胞状态还可以取一段连续的数据,如取1~10之间(包括1和10)的任意值。其次,系统演化规则是一个映射关系的有限集,而这些映射关系都是一些十分简单的演化状态迁移规则:如果一个元胞处于活着的状态,则在八个相邻元胞中有两个或三个活着时,在下一时刻继续活着,否则就死亡。另一方面,如果一个元胞处于死亡状态,但在邻域中有三个元胞处于活着的状态,则下一时刻这个元胞的状态变成活着,否则仍保持死亡状态。

这个规则模拟生命在过分拥挤或孤单时不能生存

和生命在一定条件下可以诞生的情况。可以用计算机来模拟这些规则,有趣的是,用这些简单规则定义的元胞自动机能够模拟生命活动中的生存、灭绝、竞争等复杂现象。在这个游戏中,实际上体现了一个二维的元胞自动机的运作过程。其状态迁移规则如图 38.4-1b 所示。



条件 规则号	状 态			
	(a)	(b)	(c)	(d)
状态例				
规则例				

图 38.4-1 元胞自动机的棋盘游戏模型

a)元胞自动机的棋盘游戏状态迁移例 b)状态迁移规则例

### 1.2 元胞自动机的基础

元胞自动机是一种模型,可以让大量的简单元胞在某些简单的本地规则作用下产生各种复杂的系统状态。通常,元胞自动机是一套格子的  $n$  维组合( $n$  为自然数),每个格子驻留了一个有限状态自动机,每个自动机以其相邻的,具有有限状态的元胞格的状态作为输入,然后输出一个处于同一有限状态集合的状态,它可以作如下表示:

$$\{s_1, s_2, \dots, s_n\}^{\Sigma} \rightarrow \{s_1, s_2, \dots, s_n\} \quad (38.4-1)$$

其中: $\Sigma$  表示有限状态  $s_1, s_2, \dots, s_n$  组成的子集; $\rightarrow$  为映射操作; $n$  为有限状态的个数,即元胞自动机的状态是与其相邻的自动机状态交互作用的结果。元胞自动机的相邻元胞取舍情况一般如图 38.4-2 所示。

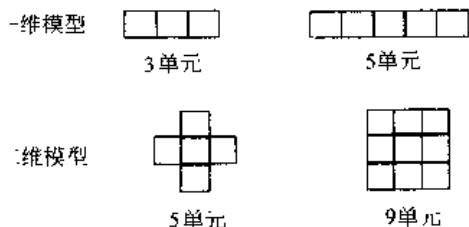


图 38.4-2 元胞和邻居元胞

以上模型均将中央元胞作为邻居,实际上,也可以不包括中央元胞。元胞自动机模型是建立在一个简单状态集和一套本地交互规则基础上的,是可以自我发展,自我完善的,可以不断扩展并可以按一定规则描述复杂系统状态和预测系统未来的自动机器。将上面“生

命的游戏”例子的元胞自动机系统数学模型化后,我们就可以用这样一个公式来表示元胞自动机模型的系统状态和规则应用情况。任意时刻某元胞的取值  $a$  由下式确定:

$$a_{i,j}^t = f(a_{i,j}^{t-1}, a_{i-1,j}^{t-1}, a_{i+1,j}^{t-1}, a_{i,j-1}^{t-1}, a_{i,j+1}^{t-1}, a_{i-1,j-1}^{t-1}, a_{i+1,j-1}^{t-1}, a_{i-1,j+1}^{t-1}, a_{i+1,j+1}^{t-1}) \quad (38.4-2)$$

事实上,这个模型是一个二维元胞自动机模型,在元胞自动机思想的应用中还可以建立一维、三维和多维元胞自动机模型。它们都是按照这样一个简单的方法建立起来的可以模拟和仿真人型复杂的静态、动态和混合形式的系统。从算术的关系角度来看,元胞自动机模型实际上是从有限状态机(Finite State Machine)演绎而来的,但是元胞自动机模型在更高层次上处理一维、二维或多维复杂系统的状态演化。它主要处理元胞阵列上元胞状态的读、写和更新的过程。

#### (1) 一维简单元胞自动机

设在一维直线上均匀地分布着  $N$  个元胞,任一时刻每个元胞可取  $k$  个整数值中的一个,某个元胞  $i$  在下一时刻的取值由现在时刻在半径为  $r$  ( $r$  为整数)内的邻居的值共同确定。用公式表示为

$$a_i^{(t+1)} = \phi(a_{i-r}^{(t)}, a_{i-r+1}^{(t)}, \dots, a_i^{(t)}, \dots, a_{i+r-1}^{(t)}, a_{i+r}^{(t)}) \quad (38.4-3)$$

其中  $a_i^{(t)}$  表示元胞  $i$  在  $t$  时刻取值,  $a_i^{(t+1)}$  是该元胞在  $t+1$  时刻的取值,  $\phi$  为某种函数关系。在最简单的情况下,  $k=2$ ,  $a_i$  可取 0,1 两个值,  $r=1$ , 式(38.4-3)简化为

$$a_i^{(t+1)} = \phi(a_{i-1}^{(t)}, a_i^{(t)}, a_{i+1}^{(t)}) \quad (38.4-4)$$

式(38.4-4)即为一维二值三邻居简单元胞自动机的演化公式。

由此我们看到,一维简单元胞自动机具有两个最显著的特点:

- 1) 元胞之间的相互作用是最邻近的局域作用。
- 2) 元胞只取 0,1 两个值,任何时刻只能取其中的一个。

对于一维二值三邻居元胞自动机来说,式(38.4-4)共有 256 组,每组给出元胞 8 三邻居八种可能组态 111,110,101,100,011,010,001,000 在下一时刻所决定的元胞  $i$  的取值,S. Wolfram 用规则号来说明不同的演化规则,如对于 90 规则(90 的二进制表示为 01011010),其八种组态的取值方法如下:

111	110	101	100	011	010	001	000
0	1	0	1	1	0	1	0

一维空间中(有限或无限),所有的元胞都按给定的规则同步更新,给定直线上所有元胞一个初始的取值分布,在某种规则演化下,元胞取值将按一定规律进

行变化,图 38-4-3 给出  $N=63$ ,初始分布只有  $i=32$  的元胞取 1,其他取 0 值,图 a 经 90 规则演化 31 步,图 b 经 150 规则演化 31 步后,将所有的分布图按时间先后从上到下排列所得到的演化图,图 38-4-4 的初始分布中,每个元胞取值 0 或 1 是随机的,其中取值 1 用黑点表示,取值 0 用空白表示。

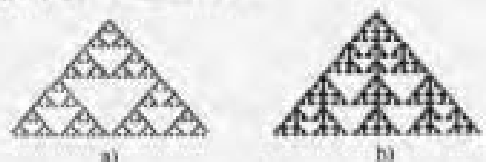


图 38-4-3 “1”值初始位形的元胞自动机演化  
a)90(010110101) b)150(100101101)

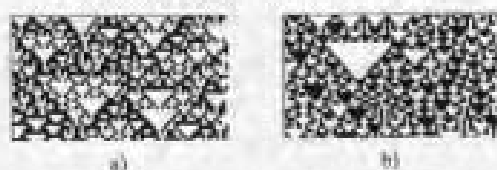


图 38-4-4 随机初始位形的元胞自动机演化  
a)90(010110101) b)150(100101101)

除一维元胞自动机外,还可以研究分布在半无限直线上的元胞自动机,这时将它的一个端点记为  $i=0$ ,而用所有非负整数表示元胞位置,在局部映射中,自变量只出现  $a_i, a_{i+1}, \dots, a_{i+r}$ ,我们称这样的元胞自动机为单向(one-way)元胞自动机。

如果局部映射的形式为

$$a_i^{t+1} = f(a_{i-r}^t + \dots + a_i^t + \dots + a_{i+r}^t)$$

即在邻域中每个元胞的状态都起同样的作用,则称为完全(Totalistic)元胞自动化。在用元胞自动机模拟某些物理现象时,往往会对局部映射提出某些附加限制,常用的有:

① 如将符号 0 看做静止状态,则要求

$$f(0, 0, \dots, 0) = 0$$

② 除了  $f$  与  $i$  无关这个反映空间齐性的条件之外,还要求  $f$  是对称的,即

$$f(a_{i-r}, \dots, a_i, \dots, a_{i+r}) = f(a_{i+r}, \dots, a_i, \dots, a_{i-r})$$

它是空间各向同性的反映。

称满足这两个条件的元胞自动机为合法(Legal)元胞自动机。

(2) 元胞自动机演化位形与形式语言描述

1) 元胞自动机演化位形的复杂性 图 38-4-5a, b, c 是一维二值三邻居元胞自动机 128, 36, 90 规则在某个范围内的随机初始位形上经 13 步演化得到的演化图,图 d 是由一维二值五邻居 52 规则演化得到。可见,一维元胞自动机演化产生的位形图可分成四类:① 随时间演化消失;② 到达一个固定的有限尺寸;③ 以一定的速率不断生长;④ 无规则地收缩和生长。前三类分

别对应于连续动力学系统的极限点,极限环和混沌吸引子。第 4 类表现复杂,人们猜想能进行通用计算。

2) 元胞自动机位形集合与形式语言 一堆  $N$  元胞元胞自动机在时刻  $t$ ,位置  $i$  取值  $a_i^t$  为  $S = \{0, 1, \dots, k\}$  中的一个,这些值组成的所有可能的序列构成元胞自动机位形的有限集合  $\Sigma S^N$ ,局部规则(1,2)完成  $2r+1$  个  $S$  中的值到一个  $S$  中值的映射,并从整体上导致元胞自动机集之内的一个映射:

$$\Phi: \Sigma \rightarrow \Sigma \quad (38-4-5)$$

一般地有

$$Q^{t+1} = \Phi Q^t \subset Q^t \quad (38-4-6)$$

其中

$$Q^t = \Phi^t \Sigma$$

在此,元胞自动机演化位形集合可看作一种形式语言, $S$  是字母集,一个元胞自动机位形等同于语言中的一个单字,演化规则中即为产生规则。

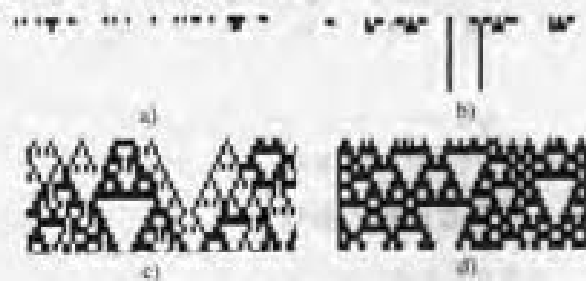


图 38-4-5 元胞自动机演化时空结构的四种类型

a)128(10000000) b)36(00100100)  
c)90(01011010) d)52(00110100)

3) 90 规则与 150 规则的形式语言 考虑一维简单元胞自动机 90 规则,从初始位形集  $Q(0) = \Sigma$  经一步演化得到  $Q(1)$ ,90 的二进制展开为 01011010,演化规则如下:

$$\begin{aligned} 111 &\rightarrow 0, 110 \rightarrow 1, 101 \rightarrow 0, 100 \rightarrow 1, \\ 011 &\rightarrow 1, 010 \rightarrow 0, 001 \rightarrow 1, 000 \rightarrow 0 \end{aligned} \quad (38-4-7)$$

对于某个初始位形  $A^{(0)} \in Q^{(0)}$ ,经一步演化所得位形  $A^{(1)} = \Phi A^{(0)} \in Q^{(1)}$ ,位形  $A^{(1)}$  中某点  $i$  的值  $a_i^{(1)}$  由位形  $A^{(0)}$  中的三点  $\{a_{i-1}^{(0)}, a_i^{(0)}, a_{i+1}^{(0)}\}$  确定,这种确定关系及  $A^{(1)}$  中各点的取值可由 Bruijn 图表示:

图 38-4-6 中各节点代表  $A^{(0)}$  中的  $\{a_{i-1}^{(0)}, a_i^{(0)}\}$ ,用弧线连接起来的两节点代表三邻居  $\{a_{i-1}^{(0)}, a_i^{(0)}, a_{i+1}^{(0)}\}$ ,弧线上的表达式即为式(38-4-7),每条弧线与一个符号值相连,图中每条可能的路径都对应一个特殊的初始位形  $A^{(0)}$ , $A^{(1)}$  则由  $A^{(0)}$  相应的路径各弧线上的符号序列组成,图中所有的路径构成了  $Q^{(1)} - Q^{(0)}$ ,图 38-4-6 即是产生形式语言  $Q^{(1)}$  的有限自动机的状态转移图,图中节点代表有限自动机的状态,每条弧代表有限自动机所接受的语言中的产生规则,若以  $a_0, a_1, a_2, a_3$  分

别对应节点 00,01,10,11,则该语言产生式为

$$\begin{aligned} u_7 &\rightarrow 0u_1, u_6 \rightarrow 1u_1, u_1 \rightarrow 0u_2, u_1 \rightarrow 1u_3 \\ u_2 &\rightarrow 0u_1, u_2 \rightarrow 1u_0, u_3 \rightarrow 0u_3, u_3 \rightarrow 1u_2 \end{aligned} \quad (38.4-8)$$

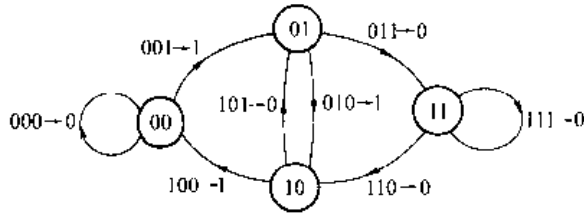


图 38.4-6 元胞自动机 90 规则的 Bruijn 图

由此得出结论:90 规则对应的语言是正则语言。

对所有的第 1、2 类一维元胞自动机,所对应的形式语言都是正则语言,但大部分第 3、4 类一维元胞自动机,对应的正则语言的复杂性(用状态图中的节点数表示)随时间迅速地增加,可能导致非正则语言极限集。对某些第 4 类元胞自动机来说,要确定某特殊的位形是否产生过则是 NP 完全性问题,90 规则是第 3 类元胞自动机,但其形式语言的复杂性却具有最简单的值,可以作比较精确的讨论。

图 38.4-6 中每个节点都有两条入线 0 和 1,两条出线 0 和 1,因此 4 个节点是等价的,其最小状态图如图 38.4-7 所示。

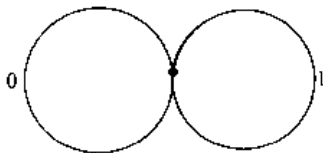


图 38.4-7 最小状态图

由图 38.4-7 可知,90 规则产生位形的正则表达式为

$$\Omega^{(1)} = ((0^*)(1^*))^* \quad (38.4-9)$$

图 38.4-7 和式(38.4-9)表明,对于任意时刻  $t$ ,90 规则产生位形的正则表达式为

$$\begin{aligned} q_0 \ a_L \ a_L \ r q_1 & \quad q_2 \ 1 \ 1 \ L q_4 & \quad q_5 \ 1 \ 1 \ r q_6 & \quad q_5 \ 0 \ 0 \ L q_4 & \quad q_7 \ 1 \ 1 \ L q_7 \\ q_1 \ 1 \ 1 \ r q_4 & \quad q_2 \ 0 \ 0 \ L q_3 & \quad q_5 \ 0 \ 0 \ r q_6 & \quad q_5 \ 1 \ 1 \ L q_7 & \quad q_7 \ 0 \ 0 \ L q_7 \\ q_1 \ 0 \ 0 \ r q_2 & \quad q_3 \ 0 \ 1 \ r q_5 & \quad q_5 \ a_R \ a_R \ L q_7 & \quad q_5 \ a_R \ a_R \ L q_7 & \quad q_7 \ a_L \ a_L \ h q_7 \\ q_1 \ a_R \ a_R \ r q_7 & \quad q_3 \ 1 \ 1 \ r q_1 & & & \\ & \quad q_4 \ 0 \ 0 \ r q_5 & & & \\ & \quad q_5 \ 1 \ 0 \ r q_1 & & & \end{aligned}$$

上面的  $T$  程序对任何初始位形计算一次的结果等同于 90 规则对该初始位形演化一步。这种计算可用 Turing 机状态转移图表示,如图 38.4-9 所示。

对于通用图灵机 UT(Universal Turing),不但可以输入某些待计算的初始位形  $S$ ,也可以输入某个具

$$\Omega^{(1)} = (0^*1^*)^* \quad (38.4-10)$$

150 规则对应的状态转移图如图 38.4-8 所示,其形式语言与 90 规则具有完全相同的形式。

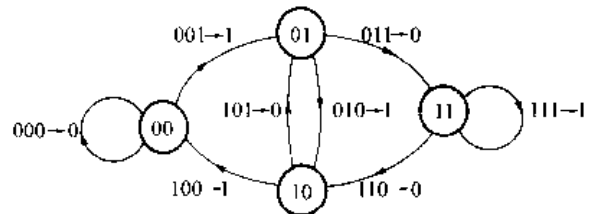


图 38.4-8 元胞自动机 150 规则的 Bruijn 图

### (3) 90 规则元胞自动机与 Turing 机

Turing 机被证明能描述任何算法。下面我们来构造 90 规则元胞自动机的 Turing 机。设一维元胞自动机在有限元胞上演化。元胞个数为  $N$ ,边界条件为恒“1”值(也可讨论其它边界条件)。设三元组基本 Turing 机为  $A = \{0, 1, a_L, a_R\}$ ,  $Q = \{q_0, q_1, \dots, q_7, q_F\}$ ,  $V = \{L, r, h\}$ ,其中  $A$  表示字母集,4 个字母中  $a_L$  作为“1”值左边界条件, $a_R$  作为“1”值右边界条件, $Q$  为状态集合,9 个内部状态中, $q_0$  是起始状态, $q_F$  是终止状态。 $v$  为动作集, $L$  表示读写头左移一格, $r$  表示右移一格, $h$  表示读写头不动。Turing 机带上格子个数可取  $N+2$ ,各格子的计算顺序从左到右,当把  $N$  个格子处理完毕后读写头重新回到左端的起始处,这一过程可写作:

$$\begin{aligned} \downarrow q_0 & \quad \downarrow q_F \\ a_L S a_R & \Rightarrow a_L S' a_R \end{aligned} \quad (38.4-11)$$

其中  $S, S'$  分别代表原位形与新位形, $\Rightarrow$  代表所有的计算过程,处理完  $N$  个格子所需要的步数为  $4N+3$ 。若要继续进行下一次计算只需将  $q_F$  置为  $q_0$ 。Turing 指令形式取为

$$q_i \ a_i \ a_i' \ v \ q_i' \quad (38.4-12)$$

其中  $a_i' \in A, q_i' \in Q, v \in V$ 。

模拟恒“1”值左、右边界的 90 规则元胞自动机演化的 Turing 机,可取以下 19 条指令:

体的  $T$  程序,若把  $S$  对应数  $x, T$  程序对应 Godle 数  $t$ (如 90 规则的编号数 90),则 UT 的输入是数组  $\langle t, x \rangle$ ,UT 对这个数组的计算等同于 Turing 机对输入  $S$  的计算。

以上我们看到,通用 Turing 机事实上能模拟出所

有的元胞自动机演化,反过来某个元胞自动机也可模拟某个特定的 Turing 机,以并行的方式在一个时间步中完成 Turing 机在若干时步所作的运算,因而降低了运算的时间复杂性,但并不是每个元胞自动机都可模拟通用 Turing 机,在一维二状态三邻居基本元胞自动机里没有哪个规则能与 U<sup>T</sup> 应。A. R. Smith 曾构造

了一个 18 状态三邻居一维元胞自动机,可等价于已知的最简单通用 Turing 机(即 Minsky 的 7 状态 4 符号机)。二维元胞自动机的“生命”(2 状态 9 邻居)游戏亦被证明是计算通用的。人们推测第 4 类元胞自动机都能进行通用计算。给定合适的编码,这些元胞自动机原则上可模拟任何其他系统,从而展示任意复杂的行为。

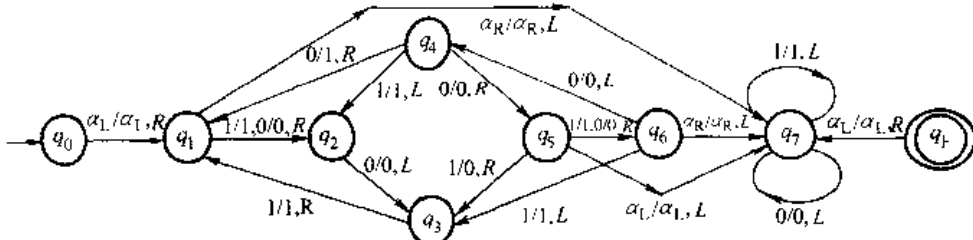


图 38.4-9 模拟元胞自动机 90 的 Turing 机状态转移图

### 1.3 元胞自动机的自组织建模方法

元胞自动机是不可判定的,换言之,不能用有限的程序步骤对元胞自动机演化图形的终态给出一般性的答案。但是元胞自动机具有强大的计算功能,它的并行运算方式为自组织建模展示了美好的前景。

#### (1) 对象的元胞自动机抽象化模型分析

元胞自动机方法是一种用来分析一个静态或动态系统的方法,它主要通过对一个系统的抽象化、模型化来定义一个系统的状态。而在此过程中,它不像一般的数学物理方法那样死板的描述或仿真一个系统。用数学物理方法分析一个较为复杂的系统时,可能是十分困难的,甚至是不可能的。而元胞自动机方法采用了一些新的思想来简化分析过程,同时可以达到较好的仿真效果,同时可能成为稳定和可靠的方法。用元胞自动机方法分析一个系统包括以下几个方面内容:

#### 1) 确定要研究的对象系统的性质

① 此系统是否一个自组织系统(self-organization)。对于一个系统,是否自组织系统是决定在分析系统过程中如何划分模块和简化数学模型的重要参照依据之一,因此在用元胞自动化方法分析系统时,若此系统是自组织系统,即在系统内部只有本地交互因素,而无外界相关因素(不变的、固定的连接因素)。那么,此系统将表现出一种相对稳定性(robust)。这样将有利于整个系统的综合分析,而实际上,系统因存在混沌吸引子(attractors)而呈现一种混沌状态。

② 此系统是一个静态系统还是一个动态系统。确定系统是动态还是静态是涉及到系统模型的重要因素。因为静态系统和动态系统的建模是完全不同的,静态系统的内部结构是稳定的,而动态系统是不稳定的,但可能其系统特性是收敛的。

③ 此系统是一维、二维系统还是三维或多维系统。系统的状态与系统的维数关系密切,同时维数不同

元胞自动机方法的应用也不同,而且还可能包括线性维和非线性维。

2) 确定系统规则 系统规则的确定是整个系统元胞自动机分析的核心。元胞自动机在处理系统时,总是将系统细分为可以建模的元胞,这些元胞间关系的分析是较为简单的。一个元胞自动机模型,在外界条件的变化确定后,然后通过规则确定整个系统的各个元胞状态,以至整个系统各个元胞的状态,因此规则的确定及完善关系到将来系统状态的演化结果。

3) 系统结构的规则 系统在确定了状态迁移之后,就可以确定系统的逻辑结构,将这些结构进行综合和优化之后,就可以将模块化的结构投入使用。

#### (2) 对象的元胞自动机模型

A. I. Adamashii 在《辨识模糊元胞自动机》中认为:元胞自动机模型是可以被多维的具有可选整型变量值的格子,其中每个元胞可以选取一个特定的集合中的某个状态,并且这些状态均可以根据一个本地的转换功能函数做出改变,而此功能函数是由附近的其他元胞的状态作为输入的因素。这个定义也就是说一个元胞自动机模型是由一系列的元胞所组成的格子状系统,其中每一个元胞的值均属于某一个集合,而且每一个元胞的值由其附近的元胞(也可以包括本身)的前一个状态决定目前状态,依次类推,整个系统就按照此功能函数作线性或非线性的状态变化。

例如在日常生活中,有些棋类就是给出一种初始状态和一系列规则,然后,由不同的对弈者选择不同的规则根据当前棋局状态做出下一个棋局状态的预测,再作决定。当然对弈的结果只有两种可能:一方胜出或和局。在上面的例子中,我们提到两个内容:

1) 分析初始状态 在以上的例子中,我们给出的初始状态只是不同的棋子布局,这种布局将是影响以后布局的源泉。假使我们在开始时给出另一种布局状态,那么,整个结局将是完全不同的。这即说明,初始状

态是影响一个系统状态的因素,但是我们又发现即使初始状态不变,每一棋子按这些规则运行下去后,整个棋局仍是极其微妙的。

2) 分析规则 规则是我们事先给出的,用来约束系统状态的条件集合。在分析一个较为复杂的系统时,我们一般先选取一个有代表性的、较小的、不是太复杂的系统来分析,通过实验等方法来确定一系列有用的关系网(规则)。一般我们可以认为这种分析是正确的。然后将这种规则应用于那些较复杂的系统,这样分析复杂系统时就较为容易。沃尔夫勒姆(S. Wolfram)在大量的计算机实验的基础上,提出将所有元胞自动机的动力学行为归纳为四类:

① 趋于一个空间平稳的构形,即指每一个元胞处于相同状态。

② 趋于一系列简单的稳定结构或周期结构。

③ 表现出混沌的非周期行为。

④ 出现复杂的局部结构,或者说是局部性的混沌,其中有些会不规则地传播。

这种分类不是严格的数学定义。前三类行为相当于低维动力系统中常见的不动点、周期与混沌,第4类行为可以与生命系统等复杂系统中的自组织现象相比拟。而我们认为第4类行为其实是高维系统中的低维混沌,仍可以归入混沌类行为,它是复杂构形的典型体现。按照沃尔夫勒姆的观点,众多(也许所有)的元胞自动机的动力学行为可归纳成数量如此之少的几类,这已是非常有意义的发现。它反映出这种分类可能具有某种普适性,很可能有许多物理系统或生命系统可以按这样的分类方法来研究,尽管在细节上可以不同,但每一类中的行为在定性上是相同的。

### (3) 对象的元胞自动机模型构造

系统分析是对一个系统进行仿真分析的第一环节,只有处理好系统分析才能有一个较好的建模。在进行系统综合分析时,首先按系统的状态受影响的复杂程度将系统分为简单系统和复杂系统。简单系统主要指系统内部较为稳定,同时所有输入只有内部本地交互动应,而无外部输入响应的系统。复杂系统一般指包括外部复合条件的响应系统,而这类系统又可以划分为多个简单系统,这些简单系统又可以按另一种元胞自动机规划的系统,这里仅论述一下简单系统。

1) 对系统进行格状分割 这个过程的关键是将整个系统看作无数个小元胞构成,而且每个小元胞应具有可选状态。当我们取定初始状态进行分析时,一定要注意每个小元胞与别的元胞发生的相互作用,即应选定本地元胞与相互作用相邻元胞数,同时此元胞数一定要远远小于整个元胞自动机系统的元胞总数。

2) 初始状态的确定 初始状态指已经确定的、经

过划分的各元胞的演化初始值,其状态可以是布尔值或者一段连续变量值,关键是应注意其初始状态要具有代表性和一般性、全面性。当然,对于已经简化了的元胞,对其进行全面分析并不十分困难,所以应尽量考虑到初始状态本身以及其受影响后,响应的种种状态变化,否则还需要对系统进行细分,或对其状态进行分段整合。

3) 规则构造 在经过以上两个步骤以后,就可以确定整个系统的规则。在确定元胞自动机规则时,要考虑以下内容:

① 系统元胞维数 维数是确定规则的基本要素。只有确定系统细分后的维数,才能考虑从什么样的基本规则入手,才能考虑某一个元胞是受线性影响,还是受非线性或复合影响。

② 元胞响应半径 一般在用元胞自动机分析时都要确定元胞响应半径,也就是元胞对那些相邻元胞状态的刺激进行响应,它们的范围怎样,例如一维系统半径就是影响其元胞状态的前几个和后几个元胞所在的范围,而二维可以确定圆形或方形半径范围。

③ 响应及属性 在一个本地响应范围内,有可能本地元胞是多重状态响应以及多重属性叠加下作出的响应,这时需要将每一种响应和属性关系进行综合。注意在进行分析时必须全面而准确。

④ 规则的时间和空间处理 元胞自动机处理系统并做出系统状态演化时,需要考虑到时间和空间段的处理。因元胞响应与时空相关,所以最终系统演化的状态也是与时间和空间相关的。

⑤ 规则的构造源 规则如何得出是整个规则构造的核心。规则确定的正确与否,直接关系到系统状态演化结果。规则确定是一个总结和完善的过程,有很多手段可以用来确定规则,如进化、神经网络学习和组织、实验观察法等。

## 1.4 元胞自动机的应用领域

早先 J. von Neumann 将元胞自动机系统看作是一个由许多元胞格所组成的可以自我演化和配置的整体,其中每个元胞格拥有 5 个相邻的元胞,并且每个元胞的状态可以取两种。这是元胞自动机系统最早应用的雏形。随后元胞自动机思想的理论和应用逐渐扩展到许多领域,例如图案识别、生物建模以及各种物理系统、计算机并行处理等。

近年,通过对元胞自动机系统的体系研究,已经较为清晰地将元胞自动机系统划分为几种类型。例如,最早由 Wolfram 在他的经典论文里提出:应用本地元胞相邻的一维元胞自动机系统的静态机制;接着 Martinet al 应用代数多项式将一维元胞自动机系统的



特点进行了表达;后来由 Dasetal 提出了基于代数矩阵的工具的元胞自动机系统,使其行为特性更为通用。最近 20 年来,基于元胞自动机思想的各种应用得到了发展,其可归类如下:

- 1) 物理系统的仿真。如对一些生长过程的模型化进行仿真、裂变反应系统的仿真、流体力学的仿真和类孤立系统的行为仿真等。
- 2) 生物的模型化。包括可自我复制模型、生物体系及其处理、脱氧核糖核酸(DNA)序列等。
- 3) 图像处理。
- 4) 语音识别。
- 5) 分类计算及素数生成的计算。
- 6) 仿真机。
- 7) 计算机体系结构。
- 8) 自测安装(BIST, Build-In-Self-Test)结构用于伪随机性、伪彻底性、确定性图案生成和信号处理。
- 9) 简单可测试性有限状态机(FSM)的综合。
- 10) 编码校错。
- 11) 伪相连存储器。
- 12) 通用及完善的哈希函数生成。
- 13) 波段失效诊断。
- 14) P 模式乘法器。
- 15) 块状和流线型密码系统。
- 16) 断裂学和混沌学。

元胞自动机思想是一种崭新的思想,它不同与传统的处理系统的思想或方法。它从根本上开辟了一条解决各种各样不同系统、不同事物空间的状态、性质的确定、优化和预测问题,同时它的应用将所有传统上不能处理或处理起来十分困难的问题,通过一种新的方法进行稳定而又可靠的化简,从而使这些问题得以妥善的解决。

元胞自动机的应用使得很多应用传统的算法解决起来相当困难或根本无法解决的问题(如生命科学仿真难题、超大规模集成电路的测试与综合问题、大型并行计算机的运算问题以及复杂动力学系统的预测和仿真问题等)有了一个崭新的突破口。在这些问题上应用元胞自动机思想不但可以减小系统分析的复杂程度,而且提高了解决问题的效率。元胞自动机同时具有极好的泛用性和稳定性。元胞自动机是一种新的计算机算法,它主要采纳了最新的系统分析思想即系统元胞化思想作为应用基础,同时,它又将系统演化的客观规律(即在系统模型被分割成很多极小的元胞以后,每个元胞的性质和表现总是受其相邻元胞性质和表现的影响,或者说由它们所确定)融入了算法的核心。所以元胞自动机的思想在各种工程、技术领域的应用有着其他传统算法思想所无法比拟的特点和优势。

但是,元胞自动机还存在一个十分重要的问题仍未得到妥善的解决,那就是元胞自动机应用起来看似简单,而在实际应用过程中要想得到完善的元胞自动机规则却很困难。所以在元胞自动机应用过程中,确定规则是关键。传统上有以下两种确定元胞自动机规则的办法:

1) 根据已知的系统演化的过程,记录各个不同时空点上系统的状态;然后从这个状态集中将各个元胞状态的变化过程,用一些线性、非线性方程或微分方程来表达。这样,在系统演化过程中,通过时空表达方程(即状态迁移方程),可以直接计算出本地元胞的状态,最终可以得到整个系统的演化状态。

2) 已知系统初始状态和最后状态,列出所有可能的规则集。对初始状态使用规则集中某一规则进行作用,这样总会存在一个或多个规则组成的规则集合可以使系统演化到最终状态,那么,这个规则集合就是所要找的规则。

以上两种方法实际上都不是较好的解决方法,其理由为:

① 对于第一种方法,如果系统比较简单,那么,系统规则是可以记录、更新和查询由状态集综合得到的系统演化状态曲线或概率点阵得到。但是,一旦系统比较复杂时,比如系统元胞状态比较多时,应用现有的计算机工具就无法满足大量的计算需求,而且目前还没有比较完善的智能化工具来综合状态集。

② 对于第二种方法,如果系统演化时每个元胞只取较少的相邻元胞作为对其有影响的元胞,那么这种方法也是可行的。但是,同样如果想得到比较完善的规则以使将来对随机系统的预测或判断更为准确而取较多的相邻元胞,比如二维元胞自动机模型中取 8 个,每个元胞状态只取两个,那么规则数就将有  $2^{256}$  种,那么这种方法根本就无法实现。

## 2 结构拓扑的自组织进化

机械结构初始拓扑形态的创新,一直是结构自动设计的一大难题。原因是机械结构的拓扑形态在很大意义上,基本决定了结构的功能、载荷、约束、材料的配置等的适用范围,即用现有的结构、材料、力学的解析法(有限元法等)和结构优化方法,由于表达结构形态设计的自由度庞大,只能在相对结构初始形态变化较小的范围内,实现结构的拓扑形态的再设计和强度、应力分布等的校核。为此,近年来基于元胞自动机的结构拓扑形态自组织化设计的研究,在机械结构拓扑形态设计方法方面开创了一个新的方向。其基本思想是利用元胞自动机多规则组合驱动的自组织演化机制,实现大自由度结构复杂拓扑形态内部机理的表达。具体

地讲是将结构分割为结构元胞单元(以下简称为结构单元),在元胞自动机的局部规则(以下简称为规则)驱动下,实现相邻结构单元状态的自组织,通过元胞自动机的演化机制表达整个系统的状态。

用元胞自动机实现结构拓扑形态的自组织问题,关键在于确定元胞自动机的规则。为了寻找其规则,现在大多数学者采用了如图 38.4-10 所示的基于局部间接规则(以下简称为间接规则)的进化元胞自动机(Evolutionary Cellular Automata;ECA)方式驱动结构的自组织演化过程。所谓间接规则,是在元胞自动机的结构拓扑形态自组织过程中,由离散单元的力学解析(FEM 等)得到的相邻单元应力和相关系数等信息构成迁移函数作为自组织演化规则的表达。迁移函数中的相关系数作为设计变量由进化过程确定。但是,由于局部规则的局部性特征以及结构的形状、材料特性等复杂因素的影响,间接规则存在着缺乏一般性、计算量大、实际操作难度大等问题。

为此,本节讨论一种不依赖于力学解析信息的局部直接规则(以下简称为直接规则),求解结构的优化拓扑初始形态的方法,即将直接规则集合作为设计变量,通过进化过程确定直接规则群,并由此决定状态演化迁移函数(以下简称为迁移函数)的形式。为了减小设计变量数,在迁移函数中,将  $t-1, t-2$  时刻中心单元的相邻单元状态作为输入,确定  $t$  时刻中心单元的状态。依此不仅解决了元胞自动机反复演化过程中 FEM 解析的计算量问题,而且有效地实现了元胞自动机规则的一般性表达,又大大减小了直接规则的自由度。在结构质量最小化评价目标下,实现规则集合进化的同时,完成结构拓扑形态的自组织。文中以简单的平面薄板结构拓扑优化为例,显示了提出的直接规则在结构拓扑优化问题中的作用和有效性。

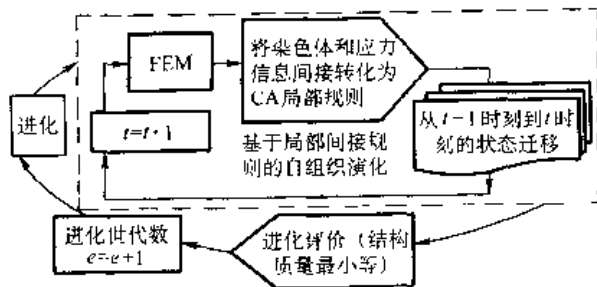


图 38.4-10 基于间接规则的 ECA 框图

## 2.1 结构拓扑优化中的 ECA 直接规则

### (1) ECA 的间接规则

在基于 ECA 间接规则的结构拓扑形态设计中,其迁移函数的确定大多是根据人们的经验或其他方法计算的结果来建立。对图 38.4-11a 所示的平面薄板质量

最小化设计时,无论采用如图 38.4-11c、d、e 的元胞自动机平面模型之一,例如可先设定如式(38.4-13)所示的迁移函数:

$$S_{ij}^e = F[\cdot] = \sum_{k=1}^n \alpha_k^{-1} \cdot \alpha_k \quad (38.4-13)$$

式中  $n$  ——如图 38.4-11c、d、e 所示的元胞自动机平面模型当前单元的相邻单元总数;

$k$  ——如图 38.4-11c、d、e 所示的元胞自动机平面模型当前单元的相邻单元序号;

$e$  ——如图 38.4-10 所示的进化世代数;

$S_{ij}^e$  —— $t$  时刻当前单元( $i, j$ )的状态;

$\alpha_k^{-1}$  —— $t-1$  时刻与  $S_{ij}^e$  相邻单元( $k$ )的应力;

$\alpha_k$  ——第  $e$  代进化的迁移函数加权系数。

使用 ECA 的间接规则进行结构拓扑优化过程如图 38.4-10 所示,用第  $e$  代进化得到的迁移函数加权系数(设计变量),以及元胞自动机自组织演化过程中  $t-1$  时刻结构拓扑形态的应力分布解析结果,根据式(38.4-13)建立的元胞自动机的间接规则,完成  $t$  时刻结构拓扑形态的状态迁移。状态迁移如此反复进行,直至  $t = T_{max}$ , (元胞自动机的最大状态迁移次数),结构拓扑形态自组织完成。再通过进化评价决定是否进入下一代迁移函数加权系数的进化。可见基于力学解析结果的间接规则,在结构的形状、材料特性等复杂因素的影响下,对于一般性结构拓扑形态设计问题难于适用。这种确定 ECA 的间接规则的方法存在以下的主要问题:

1) 基于应力信息的迁移函数数学表达形式,缺乏一般性的指导意义,事先设定的迁移函数数学表达形式与实际结构拓扑形态的力学机理无必然联系。

2) 迁移函数中的应力信息需要 FEM 解析得到,大大增加了进化解析的时间。

3) 把迁移函数中的相关系数(加权系数等)设定为实数设计变量,在进化过程中其计算精度对局部规则的影响极大。

为此,以下将以图 38.4-11a 的平面薄板设计为例,讨论一种不依赖于力学解析信息的元胞自动机直接规则的表达和基于 ECA 的结构拓扑优化设计方法。其基本思想如图 38.4-12 所示,将直接规则集合作为设计变量,在结构质量最小化和直接规则收敛性评价目标下,实现局部规则集合进化的同时,以简单的元胞自动机自组织演化机制,实现大自由度结构拓扑形态的设计。

### (2) 直接规则的构造

假如以图 38.4-11c、d、e 的元胞自动机平面模型,每个单元的状态定义为二值化,可以将结构拓扑形态问题转化为求解一个在元胞自动机的自组织演化过程

中与结构的力学解析无关, 仅与结构质量分布相关的有限直接规则集合的问题, 但是, 局部规则的定义和构造决定了其集合的大小, 也决定了设计变量数的多少。

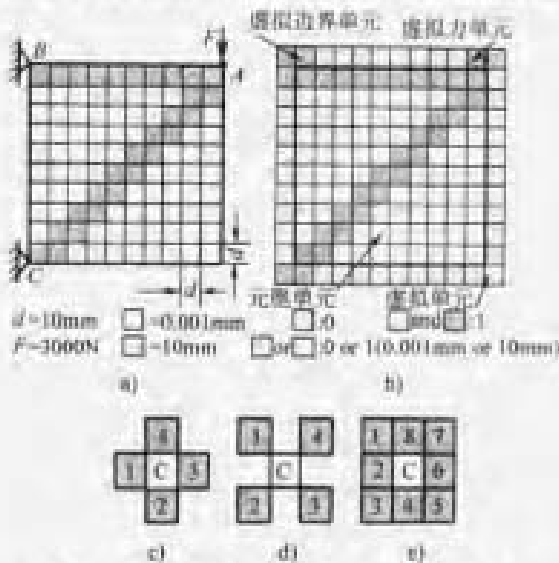


图 38-4-11 解析模型和元胞自动机的平面模型  
a) 结构模型 b) 虚拟单元 c) 元胞自动机模型 1  
d) 元胞自动机模型 2 e) 元胞自动机模型 3

按照如图 38-4-11c,d,e 所示的元胞自动机常用平面模型的类型, 图 38-4-11c 和 d 中心单元的相邻单元数较少, 难以表达复杂的结构拓扑形态。假如以图 38-4-11e 相邻单元为 8 个元胞自动机平面模型进行结构拓扑形态设计, 则其直接规则的所有组合为  $2^8=256$ 。也就是说, 其设计变量数为 256 个。这样, 相对于 ECA 的间接规则, 将大大增加其进化过程的计算压力, 为了减小进化过程的计算负担, 必须通过合理地构造直接规则的表达, 以降低局部规则集合的自由度, 减少设计变量数。为此, 这里提出采用式 (38-4-14)

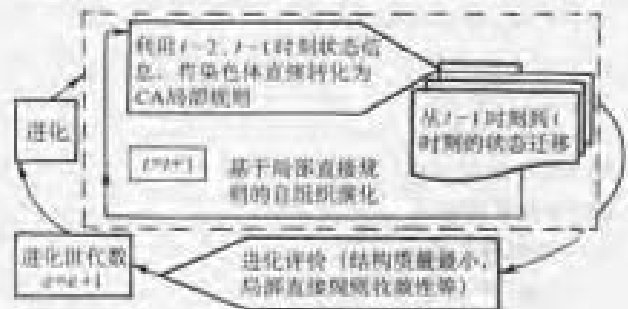


图 38-4-12 基于直接规则的 ECA 框图

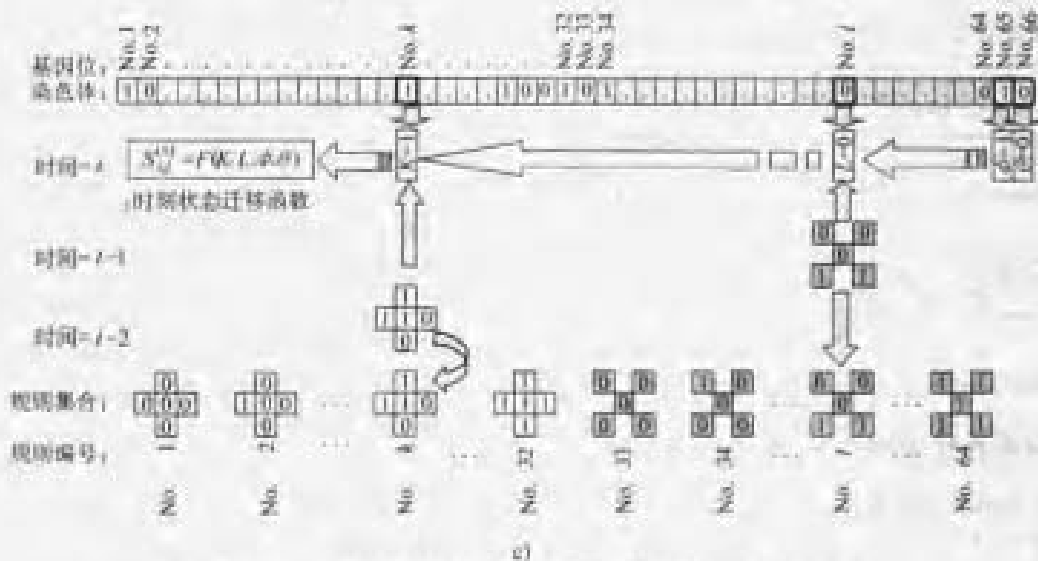
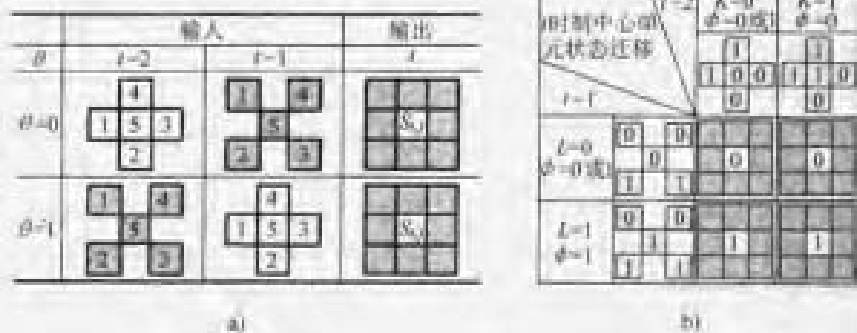


图 38-4-13 元胞自动机的直接规则, 迁移函数和遗传染色体的设定  
a) 迁移函数的选取 b) 基因 A 的作用局部规则例 ( $\theta=0$ ) c) 遗传算法染色体和元胞自动机直接规则

$$S_{i,j}^{(t+1)} = \begin{cases} F[S_{i-1,j}^{(t-2)}, S_{i,j-1}^{(t-2)}, S_{i,j}^{(t-2)}, S_{i,j+1}^{(t-2)}, S_{i+1,j}^{(t-2)}, S_{i-1,j-1}^{(t-2)}, S_{i-1,j+1}^{(t-2)}, S_{i,j-1}^{(t-1)}, S_{i,j+1}^{(t-1)}, S_{i+1,j-1}^{(t-1)}, S_{i+1,j+1}^{(t-1)}] & \theta=0 \\ F[S_{i-1,j-1}^{(t-2)}, S_{i-1,j+1}^{(t-2)}, S_{i,j}^{(t-2)}, S_{i+1,j-1}^{(t-2)}, S_{i+1,j+1}^{(t-2)}, S_{i-1,j}^{(t-1)}, S_{i,j-1}^{(t-1)}, S_{i,j+1}^{(t-1)}, S_{i+1,j}^{(t-1)}] & \theta=1 \\ S_{i,j}^{(t-1)} = S_{i,j}^{(t)} \quad (i=1,2,3; j=1,2,3) & \\ t \geq 1 & \end{cases} \quad (38.4-14)$$

所示的状态迁移函数,式中  $\theta=0$  和  $\theta=1$  两种状态转移函数的选取条件如图 38.4-13a 所示。从图中可以看出,迁移函数通过增加  $t-2$  时刻元胞自动机演化过程的状态信息,将 8 个相邻单元模型分解为两个 4 个相邻单元模型处理,使元胞自动机局部规则的组合数从 256 降为  $2^5 + 2^5 = 64$ 。 $t-2, t-1$  时刻状态所对应的图 38.4-11c、d 元胞自动机平面模型的顺序,由二值化设计变量  $\theta$  确定。

(3) 直接规则的 ECA 染色体表达

直接规则的 ECA 染色体表达,就是将以上构造的元胞自动机直接规则,在进化过程中予以染色体表达。在结构拓扑形态设计过程中,ECA 提供了搜索元胞自动机局部规则的方法,即相对于一个确定条件的结构拓扑形态设计问题,其大自由度和复杂性的难题,通过进化机制寻求能够表达问题内部力学机理的元胞自动机直接规则集合来完成,而直接规则在进化过程以染色体予以表达,这里,状态迁移函数的规则和遗传算法染色体的关系如图 38.4-13c 所示。图中,64 个二进制数的基因与 64 个直接规则相对应,同时除前述的  $\theta$  外(位于基因位 No. 66),在基因位 No. 65 再设定 1 个二进制数的基因  $\phi$ 。这样染色体由 66 个基因位来表达,其中基因  $\phi$  的作用如图 38.4-13b 所示。

图 38.4-13c 说明了在元胞自动机演化过程中,基因位 No. 1~No. 32 中  $k$  值,基因位 No. 33~No. 64 中  $l$  值,再综合图 38.4-13b 基因  $\phi$ ,最终决定了  $t$  时刻中心单元的状态。

2.2 ECA 规则的进化表达

基于 ECA 直接规则的结构拓扑形态优化设计模型,用以下公式表达:

设计变量:  $X = \{R_1, R_2, \dots, R_{61}, \phi, \theta\}^T$  (38.4-15)

目标函数:  $f(X) = \alpha W(X) + \alpha_2 Time \rightarrow \min$  (38.4-16)

约束条件:  $g(X) = |\sigma_i| - \sigma_a \leq 0$  (38.4-17)

$$h(X) = \begin{cases} Time - T_{max} \leq 0 & (t = T_{max}) \\ Time \quad t=0 & (t < T_{max}, S_{ij}^{(t-1)} = S_{ij}^{(t)}) \end{cases} \quad (38.4-18)$$

式中  $R_i$  —— 与直接局部规则相对应的二进制变量;  
 $W$  —— 结构的质量;

$\sigma_i$  —— 平面薄板内应力;

$\sigma_a$  —— 许用应力;

$Time$  —— 元胞自动机状态迁移次数;

$T_{max}$  —— 设定的元胞自动机的最大状态迁移次数;

$\alpha, \alpha_2$  —— 多目标函数的权重系数。

以上的优化模型,要求元胞自动机的状态迁移次数  $Time$  要在  $T_{max}$  以内收敛的同时,通过结构材料质量和  $Time$  的最小化目标,使系统在进化过程中保证直接规则的收敛性。通过式(38.4-15)~式(38.4-18)把基于 ECA 直接规则的结构拓扑形态优化设计问题转化为一般的多目标优化问题。

2.3 结构拓扑形态优化的算例

以图 38.4-11a 所示的问题作为算例,使用以上所提出的迁移函数和方法,讨论基于 ECA 的结构拓扑优化问题。系统的外部力学条件如图 38.4-14a 所示的四种情况,按照图 38.4-14b 所示分别将其外部力学条件转化为相应的虚拟单元,从初始元胞自动机( $e=0$ )状态开始直接规则的进化,利用元胞自动机自组织演化实现结构拓扑形态的设计。图 38.4-14c 所示为各种模型从初始状态进化到收敛状态中间结果的一例。其中进化世代数用  $e$  表示,例如图 38.4-14c 模型 4 为进化世代数  $e=300$  时状态;图 38.4-14d 模型 4 为进化的最终状态,其进化世代数  $e=429$ ,元胞自动机状态迁移次数  $Time=210$ 。

需要说明的是,以上 ECA 的进化计算以遗传算法为基础。在进化过程的  $e$  世代,由染色体所构成的直接规则集合驱动,结构拓扑形态从  $t=0$  的状态(即图 38.4-14b 的相同状态)开始迁移演化到  $t=T_{max}$ ,完成其结构拓扑形态。此时,通过 FEM 计算,在满足式(38.4-17)的应力约束和式(38.4-18)的状态迁移次数约束条件下,对染色体实现式(38.4-16)的多目标评价。从基本设计要求出发,最终得到具有收敛性特征的元胞自动机直接规则集合和结构重量最小的结构拓扑形态。

由图 38.4-14a 模型 1、2 的比较可以看出,在不同方向载荷作用下,图 38.4-14d 模型 1、2 的最终结构拓扑形态具有极其明显的最优特征;由图 38.4-14a 模型 3、4 的结构拓扑形态可见,从力学条件比较,两者的虚拟单元初始状态是相同的,但其水平方向上的受力方

向是相反的,最后得到的图 38-4-14d 模型 3.4 的结果反映了其力学模型的特征,以上,用实例较直观地证明

了提出的迁移函数和直接规则,在基于 ECA 的结构拓扑优化设计问题中的有效性。

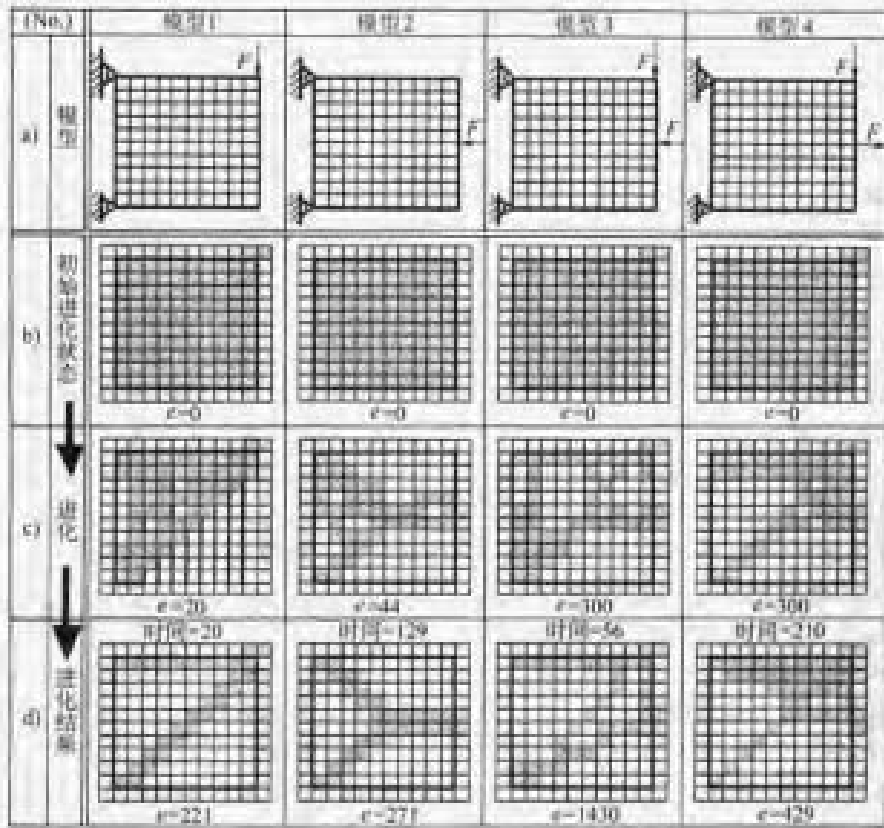


图 38-4-14 ECA 计算结果

## 第5章 自学习设计技术与方法

神经网络系统可能是我们所面临的高度复杂的非线性动力学系统,也是迄今所知功能最强、效率最高的最完善的信息处理系统,因此,很自然地成为复杂系统的建模技术。

### 1 自学习技术基础

#### 1.1 神经网络的概述

##### (1) 初创期 (1943~1969年)

1943年,心理学家 McCulloch 和数学家 Pitts 首先提出了形式神经元模型(简称 MP 模型),把神经元作为双态开关,并应用布尔逻辑的数学工具研究客观事件的形式神经网络的模拟。1948年,Wiener 在发表的《控制论》著作中,多次谈到他和 McCulloch、Pitts 以及 Rosenblatt 等人在生物神经系统、信息、控制等方面的亲切讨论、交流和合作。可以说,Wiener 的杰出工作对神经网络的研究起到了重要的推动作用。1949年,心理学家 Hebb 提出了神经元的学习规则,使神经网络具有了可塑性。到目前为止,大部分神经网络学习规则仍采用 Hebb 规则或它的改进型。1958年,Rosenblatt 提出了感知器(perceptron)模型,指出感知过程具有统计分离性,利用教师信号可以对感知器进行训练,试图模拟人脑感知能力和学习能力。1962年,Block 用解析法证明了感知器的学习收敛性;同年 Widrow 提出了自适应线性元件,它是一连续取值的神经网络,主要用于自适应系统。Minsky 和 Paper 经过对感知器模型的深入研究,在1969年发表了《感知器》一书,严格地论证了简单线性感知器功能的局限性,并且指出多层感知器还不能找到有效的计算方法。日本的中野在1969年提出一个联想记忆模型,同年,Marr 提出小脑功能及其学习法则的小脑感知机模型,这是一项神经网络与神经生理学相结合的研究成果。

##### (2) 过渡期 (1970~1986年)

自从1969年 Minsky 和 Paper 的著作发表后,曾一度使神经网络研究处于低潮,但是这项研究并未因此而中断。东京大学甘利教授从1970年起,就对神经网络的性质及其局限性作了许多理论研究,取得了一定的成果。1972年芬兰 Kohonen 发表了一个与感知器不同的线性神经网络模型,比起非线性网络模型,它的分析要容易得多。1975年福岛提出了一个自组织识别神经网络模型。1976年,美国波士顿大学教授

Grossberg 根据对生物学和心理学的研究,提出了几个非线性动力系统结构,对神经网络的研究起到了重要的推动作用。1980年,芬兰 Kohonen 提出了自组织映射理论。1982年,美国加州工学院物理学家 Hopfield 对神经网络的动态特性进行了研究,引入了能量函数的概念,给出了网络的稳定性判据,提出了用于联想记忆和优化计算的新途径。这项研究成果为神经网络的研究注入了新的活力。1984年,多伦多大学教授 Hinton 等人把神经元的输出函数与统计力学的玻耳兹曼分布联系起来,提出了 Boltzmann 机模型。1986年 Rumelhart 等人提出了误差反向传播神经网络(简称 BP 网络),它是一种能朝着满足给定的输入输出关系方向进行自组织的神经网络。这种网络实际上早在1974年前后,已被哈佛大学的 Werbos 博士所发明,只因当时没有充分体会到它的用处而多年未受到足够重视。BP 网络目前已经成为广泛使用的网络,并且以此为基础做了许多改进,发展了某些快速收敛学习算法。

从1970年至1986年间经过许多科学家长期不懈的努力和潜心研究,取得了突破性的重要成果,使得神经网络领域的研究工作摆脱了困境,并步入健康发展的新时期。

##### (3) 发展期 (1987年至今)

1987年6月在美国圣地亚哥召开了第一届世界神经网络会议,标志着神经网络研究在世界范围内形成了高潮。美国国防部预研计划管理局(DARPA)在1987年8月组织了大规模调研和论证,并于1988年11月开始一项投资数亿美元的发展神经网络及其应用研究的八年计划。此后许多国家也制定了相应计划发展神经网络。

进入20世纪90年代后,神经网络的国际会议接连不断,1990年在美国华盛顿召开;1991年在美国西雅图召开国际神经网络学术会和 IEEE 联合年会;1992年国际神经网络学会和 IEEE 神经网络委员会联合学术会议在北京召开;1993年国际神经网络会议(CNN)在美国旧金山举行;1994年首次将模糊系统、神经网络和进化计算三个方面内容的会议综合在一起召开,被称为'94IEEE 的全球计算智能大会(WCCI)在美国奥兰多召开,并决定这样的大会每三年举行一次,在这期间各学会仍单独召开年会;1995年在澳大利亚的 Perth 召开 ICNN 和 ICEC 会议;1996年在美国华盛顿召开 ICNN 会议。

1989 年我国在广州召开了全国第一届神经网络信号处理会议。1990 年在北京我国八个一级学会联合召开了神经网络首届全国学术会议。1991 年在南京召开了中国第二届神经网络学术大会。此后,于 1992、1993、1994、1995 年又分别召开了第三、四、五、六届 CCNN 大会。

总而言之,神经网络的研究无论是在国际和国内,都受到了空前的关注,并引起了许多领域研究人员的极大兴趣。尤其是进入 20 世纪 90 年代,IEEE 神经网络刊物问世以来,各种专著逐年增加,许多期刊不断推出研究专集。可以说,在世界范围内已经形成了研究神经网络的前所未有的热潮。可以预言,21 世纪初,神经网络的研究将会有更大的突破。

## 1.2 神经网络的主要特点<sup>[4]</sup>

人工神经网络是由人工神经元(简称神经元)互连组成的网络,它是从微观结构和功能上对人脑的抽象、简化,是模拟人类智能的一条重要途径,反映了人脑功能的若干基本特征,如并行信息处理、学习、联想、模式分类、记忆等。神经网络系统与现代数字计算机相比有如下不同的特点:

- 1) 以大规模模拟并行处理为主,而现代数字计算机只是串行离散符号处理。
- 2) 具有很强的鲁棒性和容错性,关于联想、概括、类比和推广,任何局部的损伤不会影响整体结果。
- 3) 具有很强的自学习能力。系统可在学习过程中不断完善自己,具有创新特点,这不同于 AI 中的专家系统,后者只是专家经验的知识库,并不能创新和发展。
- 4) 它是一个大规模自适应非线性动力系统,具有集体运算的能力,这与本质上是线性系统的现代数字计算机迥然不同。

这里,我们研究神经网络的目的是为了具有接近人脑信息处理能力的系统。因此,神经网络的样本应该是人类信息处理系统的中枢(脑)。实现这样的系统,一是为了应用,另一方面是为了搞清人脑信息处理的机理,人类对模糊信息具有非常巧妙的处理能力。让计算机具有这种能力,一直是工程技术人员和研究者的理想。但是,用传统的人工智能技术和计算机技术实现人类的日常思维活动,从本质上讲是极其困难的。

从广义角度讲,微积分、文字翻译、推理等都是——计算过程,而从数学观点看,计算就是在满足一定公理、定理的条件下,从一空间到另一空间的代数映射;从物理观点看,计算是按照一定的自然规则,在某种“硬件”上所发生的一些物理规则。因此,计算可表示为一动力系统中的状态间变换的轨迹。神经网络的计

算就是其中状态的转换,其计算过程可以认为是状态的转换过程,对给定的输入,其计算结果即是系统的稳定状态。

总之,神经网络的计算过程是适用于人类的信息处理系统。神经网络模型用于模拟人脑神经元活动的过程,其中包括对信息的加工、处理、存储和搜索等过程。它具有如下基本特点:

### (1) 具有分布式存储信息的特点

它存储信息的方式与传统的计算机的思维方式是不同的,一个信息不是存在一个地方,而是分布在不同的位置。网络的某一部分也不只存储一个信息,它的信息是分步式存储的。神经网络是用大量神经元之间的连接及对各连接权重分布表示特定的信息。因此,这种分布式存储方式即使当局部网络受损时,仍具有能够恢复原来信息的优点。

### (2) 对信息的处理及推理过程具有并行的特点

神经网络可以看作为由多数处理单元(processing element)同时动作、并行处理的机器。这里的处理单元是人工神经细胞。人脑中大约有  $140 \times 10^9$  个神经细胞进行并行处理。现在能够实现的神经网络所具有的人工神经细胞数,是  $140 \times 10^2$  个,还极其少。人脑的信息处理系统可以认为是阶层型结构,如图 38.5-1 所示的并行分散处理系统,在阶层内各模块间的相互结合呈阶层状的并行分散处理。这种并行分散处理系统,同由一个中央处理器顺次执行程序计算机是不同的,它的许多模块(单元)在相互影响的同时进行不同的处理。总之,每个神经元都可根据接受处的信息作独立的运算和处理,然后将结果传输出去,这体现了一种并行处理。神经网络对于一个特定的输入模式,通过前项计算产生一个输出模式,各个输出节点代表的逻辑概念被同时计算出来。在输出模式中,通过输出节点的比较和本身信号的强弱而得到特定的解,同时排出其余的解。这体现了神经网络并行推理的特点。

### (3) 对信息的处理具有自组织、自学习的特点

神经网络中各神经元之间的连接强度用权重的大小来表示,这些权重可以事先定出,也可以为适应周围变化的环境而不断地调整权重(自组织能力)。这种过程称为神经元的自学习过程。神经网络所具有的自学习过程模拟了人的形象思维方法,这是与传统符号逻辑完全不同的非逻辑非语言的方法。神经网络根据给定的学习数据,可以自学习。因此,不需要人类进行非常复杂的并行处理系统的编程,这可谓是一大优点。

在这里,让我们来考察一下人类大脑的学习。人类大脑的学习过程(如图 38.5-2 所示),具有进化阶段的学习、发育阶段的学习和日常学习等多重结构的学习形式。刚出生的小孩,因为已经获得了进化过程的学习

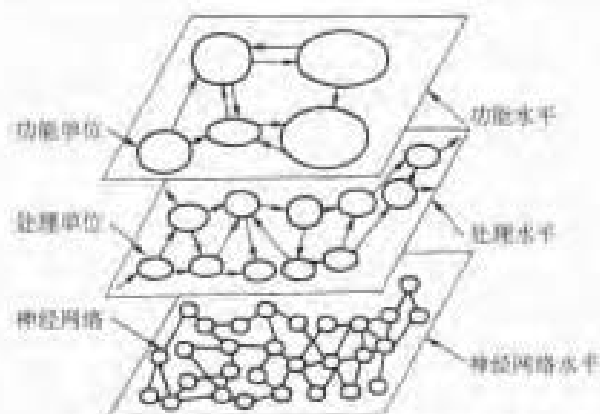


图 38.5-1 作为并行分散处理系统的脑模型



图 38.5-2 人类的多重学习结构

结果,所以具有先天的能力。在这个阶段,可以讲大脑的硬件结构已基本完全形成。在以后的发育阶段中,在

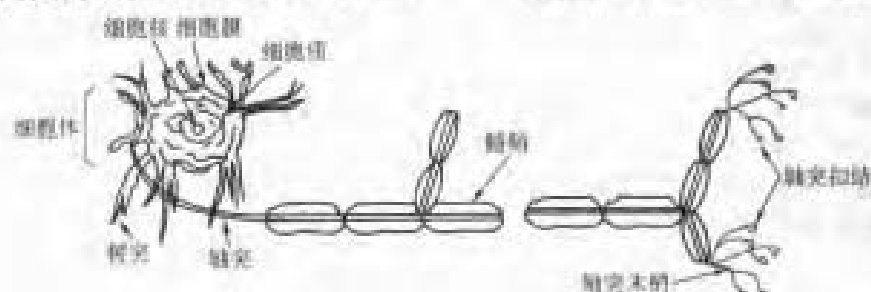


图 38.5-3 神经元的结构

外面是一层厚为  $5\sim 10\mu\text{m}$  的细胞膜,膜内有一个细胞核和细胞质。神经元的细胞膜具有选择性的通透性,因此会使细胞膜的内外液的成分保持差异,形成细胞膜内外之间有一定的电位差,这个电位差称为膜电位,其大小受细胞体输入信号强弱而变化,一般约在  $20\sim 100\text{mV}$ 。树突是由细胞体向外伸出的许多树枝状较短的突起,长约  $1\text{mm}$  左右,它用于接受周围其他神经细胞传入的神经信号。轴突是由内向外伸出的最长的一条纤维,其长度一般从数厘米到  $1\text{m}$ ,远离细胞体一侧的轴突末端有许多分支,称为轴突末梢,或称神经末梢。其上有许多扣结,称为轴突扣结。轴突通过轴突末梢向其他神经元传递信息。一个神经元的轴突末梢和

接受外部环境的影响和语言学习的同时,形成其人格,并进行基本的学习。发育阶段的学习效果,在幼儿期最显著,此阶段大约进行到 20 岁前后。一般地讲,发育阶段学习的东西,与其日常学习中学习的东西(如功课的记忆)相比较是难以忘却的,这在神经生理学中叫做学习方式的不同。

现在,神经网络的学习,还未把这样的人类学习过程完全引入,同时,人类对于人类大脑学习的机理也未完全搞清楚,因此,将来的神经网络学习方式,可以从人类大脑学习的方式得到启发,从人类大脑研究最新成果中得到发展。

### 1.3 细胞元模型

人类的智能是长期进化的结果,作为人类智能的大脑是一块很有组织的高度复杂的物质,人类大脑的神经细胞总数达 140 亿个,它是神经系统的结构和功能单元。神经细胞种类繁多,人类约有 50 多种,其大小形状也各不相同,直径在  $4\sim 150\mu\text{m}$  之间,它们在结构上具有许多共性,且在接收或产生信息,传递和处理信息方面有着相同的功能。

#### (1) 神经元的结构

神经元由细胞体、树突和轴突等组成,其结构如图 38.5-3 所示。

细胞体由细胞核、细胞质和细胞膜组成,细胞体的

另一个神经元的树突或细胞体之间,通过微小间隙相连接,这样的连接称为突触。突触的直径约为  $0.5\sim 1\mu\text{m}$ ,突触间隙有 200 个 ( $1\text{\AA} = 10^{-10}\text{m}$ ) 数量级。从信息传递过程看,一个神经元的树突,在突触处从其他神经元接受信号,这些信号可能是激励性的,也可能是抑制性的。突触有兴奋型和抑制型两种形式。

总之,至今已知人脑中有多种类的神经元存在,其中许多具有高级功能。可是现在使用的人工神经元模型是非常简单的,作为人工神经元,根据其生理学特征,通常使用图 38.5-4 所示的多输入、单输出的单元。 $n$  个神经元接受输入信号,设输入信号分别为  $x_1, x_2, \dots, x_n$ 。在第  $i$  个轴突上,单位强度信号输入时把受到



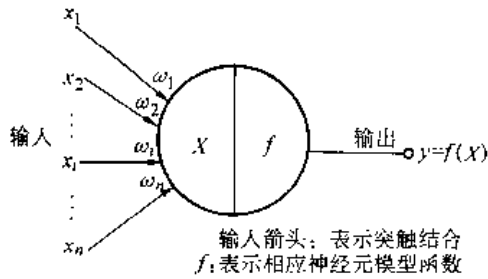


图 38.5-4 神经元形式化结构模型

影响而变化的膜电位的量用  $\omega_i$  表示。 $\omega_i$  是表示突触结合效率的量，称为突触结合强度，或叫做连接权重。对于兴奋性神经细胞的突触， $\omega_i > 0$ ；对于抑制性神经细胞的突触， $\omega_i < 0$ 。当  $\omega_i = 0$  时，可以理解为没有与第  $i$  个神经元结合。

(2) 神经元的响应特性

一般的神经元基于上述的动作功能，具有以下特性：

1) 时空整合功能

① 空间总和特性 单个神经元在同一时间可以从别的神经元接受多达上千个突触的输入，这些输入到达神经元的树突、细胞体和轴突的分布各不相同，对该神经元影响的权重也不相同。所以单个神经元对于来自空间四面八方的输入信息具有进行加工处理、空间总和的功能。其空间总和的定量描述是：膜电位的响应与输入信号和权重的线性组合有关。

② 时间总和特性 由于输入信号影响会短时间地持续，故与后到达的输入信号影响的叠加同时起作用，即神经元对于不同时间通过同一突触的输入信号具有时间总和的功能。

③ 时空整合作用 神经元把不同时间、位于不同部位的突触输入进行加工处理，决定其输出的大小的过程，称为时空整合作用。 $t$  时刻膜电位的变化定量描述为

$$\sum_{i=1}^n \int_{-\infty}^t \omega_i(t-t')x_i(t')dt' \quad (38.5-1)$$

式中  $x_i(t')$ ——第  $i$  个神经元时间  $t'$  的输入信号。

2) 阈值特性 神经元的输入输出关系具有非线性特性，如图 38.5-5 所示，即

$$y = \begin{cases} \bar{y} & \theta \geq 0 \\ 0 & \theta < 0 \end{cases} \quad (38.5-2)$$

式中  $\theta$ ——阈值。

3) 不应期 阈值  $\theta$  不是一个常数，随着神经元的兴奋而变化。绝对不应期，即无论多么强的输入信号到达，也不会输出任何输出信号的期间，可以看作  $\theta$  值上升为无穷大。

4) 疲劳 一个神经细胞持续兴奋，其阈值慢慢增加，神经细胞很难兴奋的现象叫做疲劳。

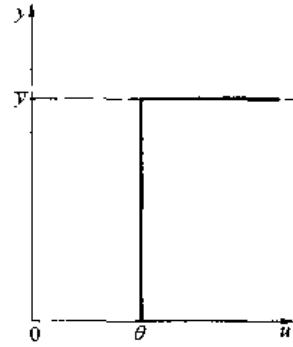


图 38.5-5 阈值特性

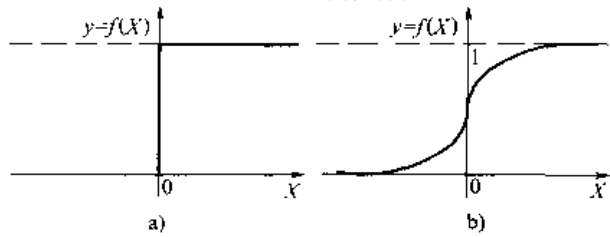


图 38.5-6 典型的神经元模型 ( $\theta=0$ )

a) 二值函数  $y = \begin{cases} 1 & X \geq 0 \\ 0 & X < 0 \end{cases}$

b) Sigmoid 函数  $y = f(x) = \frac{1}{1 + \exp(-X)}$

5) 突触结合的可塑性 突触结合强度，即权重  $\omega_i$  不是一定的，根据输入和输出信号的可塑性变化，并且可以认为由于这个变化导致具有长期记忆和学习的生理机能。

6) 神经元的模型 当不考虑时间总和只考虑空间总和时，其膜电位的变化是神经元输入信号 ( $x_i$ ) 的线性和 ( $X$ )； $X$  的响应特性由响应函数决定。响应函数模型有多种多样，以下在图 38.5-6 中列举几种典型的神经元函数模型。

① 阶跃函数模型

$$y = f(X) = U(X), \quad X = \sum_{i=1}^n \omega_i x_i - \theta \quad (38.5-3)$$

其中： $\theta$  为阈值； $U()$  为阶跃函数； $y$  为神经元的输出； $X$  为膜电位，如图 38.5-6a 所示。

这个模型是 1943 年由 McCulloch 和 Pitts 提出的，在神经网络的研究初期，被 Perceptron 神经网络所采用。

② Sigmoid 函数模型

$$y = f(X) = \frac{1}{1 + \exp(-X)}, \quad X = \sum_{i=1}^n \omega_i x_i - \theta \quad (38.5-4)$$

这个函数模型与阶跃函数模型相比，是采用了连续的信息作输入和输出，可反映实际神经方向传播网络，即著名的 BP 网络。

③ 概率函数模型。这种模型的输入和输出信号采用 0 与 1 的二值信息，它是把神经元的动作以概率状态变化的规则模型化。由输入  $X$  到输出  $y$  给出的概率分布形式为

$$p(y = 1) = \frac{1}{1 + \exp(-X/T)}, \quad X = \sum_{i=1}^n \omega_i x_i - \theta \quad (38.5-5)$$

其中： $T$  表示网络温度的正数。作为概率神经元模型应用有 Boltzmann 神经网络。

除此而外，还有时间滞后型神经元模型。

### 1.4 神经网络模型

人脑中大量的神经细胞都不是孤立的，而是通过突触形式相互联系，构成结构和功能十分复杂的神经网络系统。为了便于从结构出发模拟智能，必须将神经元连接成神经网络。其连接形式的不同，决定了神经网络的结构类别和功能。一般，按照拓扑结构，神经网络的结构大体可分为层状和网状两大类。根据人工神经网络对生物神经系统的不同组织层次和抽象层次的模拟，神经网络模型可分为：

1) 神经元层次模型：研究工作主要集中在单个神经元的动态特性和自适应特性，探索神经元对输入信息有选择的响应和某些基本存储功能的机理。

2) 组合式模型：它由数种相互补充、相互协作的神经元组成，用于完成某些特定的任务，如模式识别、机器人控制等。

3) 网络层次模型：它是由许许多多相同神经元相互连接成的网络，从整体上研究网络的集体特性。

4) 神经系统层次模型：一般由多个不同性质的神经网络构成，以模拟生物神经的更复杂或更抽象的性质，如自动识别、概念形成、全局稳定控制等。

5) 智能型模型：这是最抽象的层次，多以语言形式模拟人脑信息处理的运行、过程、算法和策略。

这些模型试图模拟如感知、思维、问题求解等基本过程。无论哪种神经网络，其神经元间的连接强度可通过学习改变。在涉及神经网络之前，先就神经元间的连

接的约束和动作给出以下约定。

#### (1) 有关神经网络模型的约定

1) 神经网络中神经元间连接的约束 人脑中约有 140 亿个神经元，不用说它们并不是全部都相互连接着的。一个神经元同其他神经元相连接的个数约 1 万个。再之，脑的功能是以其局部相对应的模块结构形式而构成的，信息从下位的神经网络层向上位的神经网络层传递，以层状结构形式而存在。因此，人工神经网络中神经元间的连接，同样受着此类约束。

2) 神经网络中神经元的同期性问题 神经网络中，各个神经元的动态动作，决定了神经网络特性的不同。因此，一般情况下，约定所有的神经元以同一时钟进行同期动作。

3) 神经网络中神经元的均一性问题 脑中存在着多种不同类别的神经元，即使是相同类别的神经元其特性也存在着差别。从实用角度讲，现在的神经网络中，其所有神经元大多以无差别、相同形式给出，而实现其功能。

#### (2) 神经网络的结构形式

1) 模块型网络 如图 38.5-7a 所示，它包含着复数个神经元，在系统中具有一定的独立功能的部分叫模块。模块的概念对理解和处理大系统是非常重要的。我们已经知道，即使在脑中也具有各种各样功能的模块，其存在于脑的特定位置。神经网络是把神经元作为基本处理器的并行处理系统。更进一步讲，也可把图 38.5-7 所示的模块作为基本处理器的并行处理系统。

2) 前向网络 如图 38.5-7b 所示的那样，神经元形成层状集团群，各神经元之间没有反馈，信号仅在层间以特定的方向传递的结构形式，叫做层次结构。也可以把单个集团看作为模块。在生物体的神经系统中，从解剖学角度看，存在着像末梢器官和大脑皮质那样的层状硬件结构；把脑看作为信息处理系统，也可认为其具有层状软件结构。

三层前向网络分为输入层、隐层和输出层。在前向网络中有计算功能的节点称为计算节点，而输入节点无计算功能。

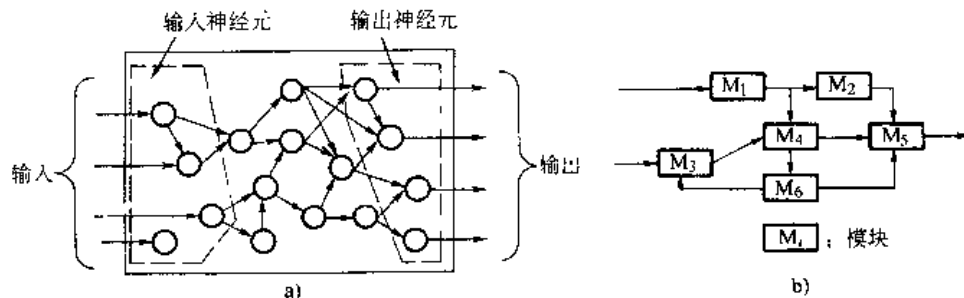


图 38.5-7 模块型网络  
a) 模块 b) 模块间的网络

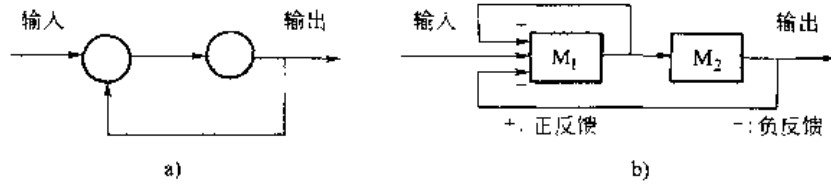


图 38.5-8 反馈

a) 神经元间的反馈 b) 模块间的反馈

3) 相结合型网络 相结合型网络如图 38.5-13a 所示,属于网状结构。构成网络的各个神经元都可能相互双向连接。所有神经元既作输入,同时也可用于输出,与前向网络不同。在网络中,如果在某一时刻从神经网络的外部施加一个输入,各个神经元一边相互作用,一边进行信息处理,直到使网络所有神经元的活跃度和输出值收敛于某个平均值为止。

4) 反馈和反馈网络 所谓反馈,从机理上讲,是将某个单元(或模块)的输出发生某种变形后的信息,再作为其系统自身的输入信息(如图 38.5-8 所示)。生物系统具有各种各样的形态,为了确保系统的稳定,反馈机理是不可缺少的。一般情况下,相结合型网络必定包含有反馈;神经网络的学习过程也必定包含着反馈评价的过程。反馈有正反馈和负反馈,正反馈具有激振机理,负反馈具有抑制控制机理。例如,在生物体中,癌的发作就是负反馈的抑制控制机构出现问题而引起的;脑中神经系统的时钟同期功能也许是正反馈激振作用的结果。

反馈网络(如图 38.5-9 所示)从输出层到输入层有反馈,即每个节点同时接受外来输入和来自其他节点的反馈输入,其中也包括神经元输出信号引回到本身输入构成的自环反馈。这里,每个节点都是一个计算单元。

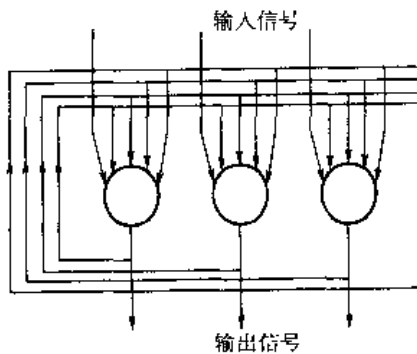


图 38.5-9 反馈网络

5) 混合型网络 混合型网络(如图 38.5-10 所示),是介于相结合型和层状结构中间的网络连接结构形式,即在前向网络的同一层中,神经元有互联的结构。这种在同一层内的互联,目的是为了限制同层内神经元同时兴奋或抑制的神经元数目,以完成特定的功

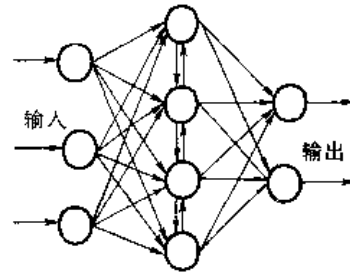


图 38.5-10 混合型网络

能。

6) 侧抑制网络 侧抑制(lateral inhibition)网络(如图 38.5-11 所示)是某个单元兴奋时,由其周围的单元实现其抑制的网络结构。侧抑制的机理主要是起到一个强调和抽出外部输入信号的特征的作用。例如,在人类的视觉系统中,对图像边缘的强调作用,就是使视觉神经产生侧抑制的效果。

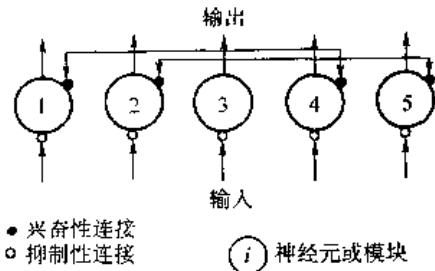


图 38.5-11 侧抑制网络

### 1.5 神经网络的学习

神经网络学习的本质特征,在于神经细胞特殊的突触结构所具有的可塑性连接,而如何调整连接权重就构成了不同的学习算法。因此,所谓神经网络的学习,就是相对于神经网络信息处理的目的,进行有机地调节其系统中神经元间的权重。如图 38.5-12 所示,人类通过感觉器官接受外界的刺激,再通过控制肌肉或器官,作用于外界。同外界的信息交流对于大脑的学习是非常重要的。例如,在幼儿学习发音的过程中,通过听觉听到母亲的声音,试着模拟其声音。然后再反馈自己的声音到自己的听觉同母亲的声音进行比较,判断自己的发音是否正确。

外界总是在变化的。因此,人类为了适应外界,总是不断地去学习。原则上,神经网络与人类相同,有必

要通过学习进行调整,适应外界的变化。以下对神经网络学习的要点作一简述:

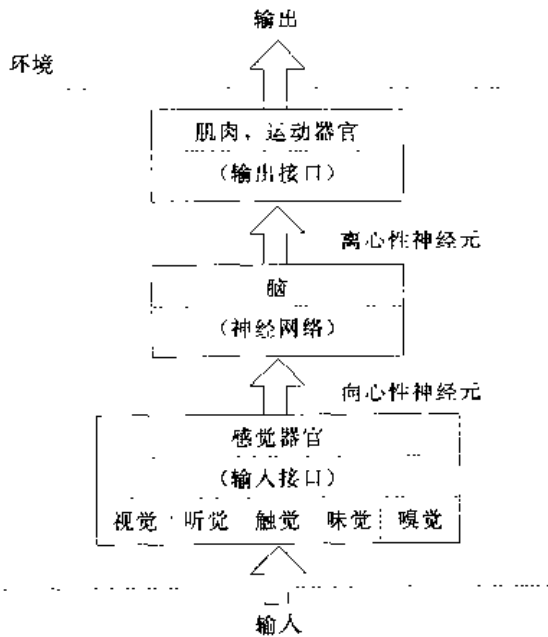


图 38.5-12 脑和外界的关系

1) 目的性的学习 生物体内存在着各种各样的学习机理,这可以说是在长期进化过程中获得的。这些仿佛具有某种目的性的巧妙机理,常常使我们不可思议。可以认为在我们的脑中也有着各种各样的目的性学习系统。

神经网络也同样,为了实现某种目的,需要为其准备学习算法。因此,为了评价神经网络的性能,需要设定评价标准,去评价其是否满足目的,或满足目的的程度。

2) 权重矩阵的调整 神经网络中神经元间的权重可以用矩阵的形式表示。例如,图 38.5-13a 的相互结合型网络的权重矩阵可用表 38.5-1 表示。其次,图 38.5-13b 的层状型网络的权重矩阵,就是将表 38.5-1 对角线部分除去后的部分。因此,可以理解为层状型网络是相互结合型网络的特殊形式。

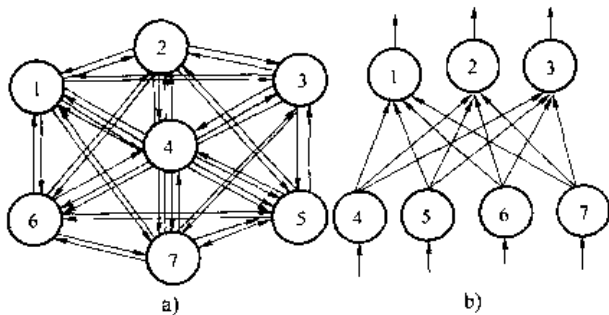


图 38.5-13 典型的网络结构

a) 相互结合型网络 b) 层状型网络

3) 学习的过程 神经网络的学习,一般以图 38.5-14 所示的过程进行。首先,设定初始权重。如果没有初始权重的设定知识,可以随机值设定。其次,输入学习数据,相对于权重参照评价标准进行评价。然后,根据评价标准调整权重,再进行评价。此过程反复进行,逐渐接近最优值。

表 38.5-1 权重矩阵

	1	2	3	4	5	6	7
1	$\omega_{11}$	$\omega_{12}$	$\omega_{13}$	$\omega_{14}$	$\omega_{15}$	$\omega_{16}$	$\omega_{17}$
2	$\omega_{21}$	$\omega_{22}$	$\omega_{23}$	$\omega_{24}$	$\omega_{25}$	$\omega_{26}$	$\omega_{27}$
3	$\omega_{31}$	$\omega_{32}$	$\omega_{33}$	$\omega_{34}$	$\omega_{35}$	$\omega_{36}$	$\omega_{37}$
4	$\omega_{41}$	$\omega_{42}$	$\omega_{43}$	$\omega_{44}$	$\omega_{45}$	$\omega_{46}$	$\omega_{47}$
5	$\omega_{51}$	$\omega_{52}$	$\omega_{53}$	$\omega_{54}$	$\omega_{55}$	$\omega_{56}$	$\omega_{57}$
6	$\omega_{61}$	$\omega_{62}$	$\omega_{63}$	$\omega_{64}$	$\omega_{65}$	$\omega_{66}$	$\omega_{67}$
7	$\omega_{71}$	$\omega_{72}$	$\omega_{73}$	$\omega_{74}$	$\omega_{75}$	$\omega_{76}$	$\omega_{77}$

4) 全部搜索法 (权重调整的问题) 在此,以学习方法中的全部搜索法 (exhaustive search method) 说明权重调整的方法。

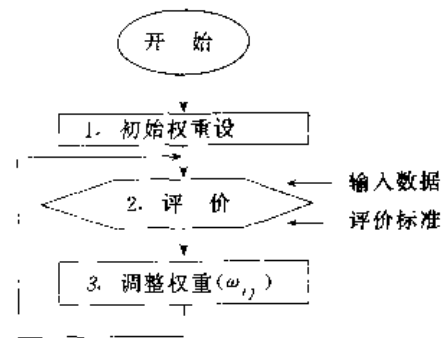


图 38.5-14 学习的过程

为了简单说明起见,首先假定把每个权重  $\omega_{ij}$  离散为  $K$  个值,应考虑的连接数为  $M$ ,则可选权重  $\omega_{ij}$  的组合数为  $K^M$  个。那么,通过学习数据和评价标准进行评价,求解最优权重值的组合。从这个意义上讲的学习,虽然在原理上是非常简单的,可是假如连接数  $M$  较大,其组合数将相当庞大。例如,即使是  $K=10, M=100$  的小规模网络,必须进行  $K^M$  次评价,这个计算量是难以实现的。因此,需要高效率的学习方法。

神经网络按学习的方式分为有教师学习和无教师学习两大类,图 38.5-15 给出了其直观描述。

(1) 有教师学习

为了使神经网络在实际应用中解决各种问题,必须对它进行训练,就是从应用环境中选出一些样本数据,通过不断地调整权重矩阵,直到得到合适的输入输出关系为止,这个过程就是对神经网络的训练过程。这种训练过程需要有教师的示教,提供训练数据,又称为样本数据。在训练过程中又需要教师的监督,故这种有教师的学习又称为监督式学习。

如图 38.5-15 所示,在有教师学习中所提供的样

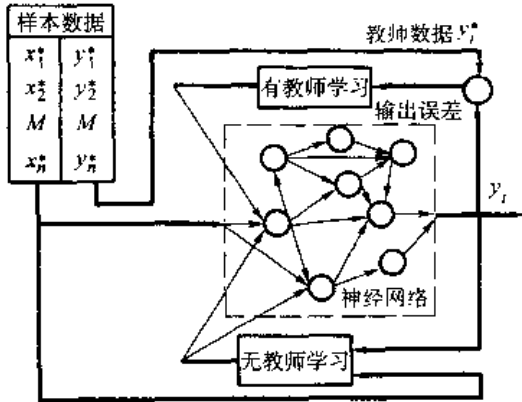


图 38.5-15 神经网络有、无教师学习直观示意图

本数据集,是指成对的输入  $x_i^*$  和输出  $y_i^*$ ; 这样的样本数据集  $\{x_i^*, y_i^*\}$  实际上代表了实际问题输入输出的关系。训练过程就是根据网络输入的  $x_i^*$  和网络输出的  $y_i$  误差来调整权重,调整是基于奖惩式规则。示教者提供正确答案,当网络回答正确时,调整权重朝着强化正确(即奖励)的方向变化;当网络响应错误时,调整权重往弱化错误(即惩罚)方向变化。

有教师学习方法虽然简单,但是要求教师对环境和网络的结构比较熟悉,当系统复杂、环境变化时,就变得困难。为了适应环境变化就要重新调整权重。这样,学习到新知识的同时,也容易忘掉已学过的旧知识。这些是有教师学习方法的缺点。

(2) 无教师学习

无教师学习的训练数据集中,只有输入而没有目标输出,训练过程神经网络自动地将各输入数据的特征提取出来,并将其分成若干类。经过训练好的网络能够识别训练数据集以外的新的输入类别,并响应获得不同的输出。显然,无教师的训练方式可使网络具有自组织和自学习的功能。

(3) 学习规则

1) 联想式学习(Hebb 规则) 心理学家 Hebb 基于对生理学和心理学的研究,在 1949 年提出了学习行为的突触联系和神经群理论;突触前和突触后两个同时兴奋(即活性度高,或处于激发状态)的神经元之间的连接强度将得到增强。后来,这一思想被其他研究者加以引用,并以多种形式加以数学表达,称为 Hebb 规则。

如图 38.5-16 所示,从神经元  $u_j$  到神经元  $u_i$  的连接强度,即权重的变化  $\Delta\omega_{ij}$  可用下式表达:

$$\Delta\omega_{ij} = G[a_i, y_i^*(t)] \times H[\bar{y}_j(t), \omega_{ij}] \quad (38.5-6)$$

式中  $y_i^*(t)$  —— 神经元  $u_i$  的教师信号;

$G$  —— 神经元  $u_i$  的活性度  $a_i(t)$  和教师信号  $y_i^*(t)$  的函数;

$H$  —— 神经元  $u_i$  输出  $\bar{y}_i(t)$  和连接权重  $\omega_{ij}$  的函数。

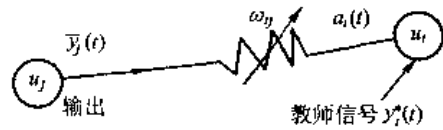


图 38.5-16 Hebb 学习规则

输出  $\bar{y}_j(t)$  与活性度  $a_i(t)$  之间满足如下关系式:

$$\bar{y}_j(t) = f_j[a_i(t)] \quad (38.5-7)$$

式中  $f_j$  —— 非线性函数。

当上述的教师信号  $\bar{y}_j(t)$  没有给出时,函数  $H$  只与输出  $a_i(t)$  成正比,于是式(38.5-6)可变为以下更为简单的形式:

$$\Delta\omega_{ij} = \eta a_i \bar{y}_j \quad (38.5-8)$$

式中  $\eta$  —— 学习率常数( $\eta > 0$ )。

上式表明,对于一个神经元较大的输入或神经元活性度大的情况,它们之间的连接权重会更大。

Hebb 学习规则的哲学基础是联想,在这个规则基础上发展了许多非监督式联想学习模型,依据确定的学习算法自行调整权重,其数学基础是输入和输出间的某种相关计算。因此,Hebb 学习规则又称为相关学习,或并联学习。

2) 误差传播式学习(Delta 学习规则) 前述的函数  $G$  与教师信号  $y_i^*(t)$  和神经元  $u_i$  实际的活性度  $a_i(t)$  的差值成比例,即

$$G[a_i(t), y_i^*(t)] = \eta_1 [y_i^*(t) - a_i(t)] \quad (38.5-9)$$

其中:  $\eta_1$  为正数,把差值  $[y_i^*(t) - a_i(t)]$  称为  $\delta$ 。

此外,函数  $H$  与神经元  $u_j$  的输出  $\bar{y}_j(t)$  成比例,即

$$H[\bar{y}_j(t), \omega_{ij}] = \eta_2 \bar{y}_j(t) \quad (38.5-10)$$

其中:  $\eta_2$  为正数。

根据 Hebb 学习规则可得

$$\begin{aligned} \Delta\omega_{ij} &= G[a_i, y_i^*(t)] \times H[\bar{y}_j(t), \omega_{ij}] \\ &= \eta_1 [y_i^*(t) - a_i(t)] \cdot \eta_2 \bar{y}_j(t) \\ &= \eta [y_i^*(t) - a_i(t)] \cdot \bar{y}_j(t) \end{aligned} \quad (38.5-11)$$

式中  $\eta$  —— 学习率常数( $\eta > 0$ )。

在上式中,如果将教师信号  $y_i^*(t)$  作为期望输出  $d_i$ ,把  $a_i(t)$  理解为实际输出  $y_i$ ,则上式变为

$$\Delta\omega_{ij} = \eta (d_i - y_i) \bar{y}_j(t) = \eta \cdot \delta \cdot \bar{y}_j(t) \quad (38.5-12)$$

其中:  $\delta = d_i - y_i$  为期望输出与实际输出的差值。因此,称上式为  $\delta$  规则,或误差修正规则。

根据这个规则的学习算法,通过反复迭代运算,直至求出最佳的  $\omega_{ij}$  值,使  $\delta$  达到最小。

上述  $\delta$  规则只适用于线性可分函数, 不适用于多层网络非线性可分函数。广义的规则将在后述的 BP 算法中详细介绍。

3) 概率式学习 从统计力学、分子热力学和概率论中关于系统稳态能量的标准出发, 进行神经网络学习的方式, 称为概率式学习。概率式学习的典型代表是 Boltzmann 机学习规则, 它基于模拟退火的统计优化方法, 因此又称为模拟退火算法。

Boltzmann 机模型是一包括输入、输出和隐层的多层网络, 但隐层间存在着互联结构且网络层次不明显。对于这种网络的训练过程, 就是根据下述规则对神经元  $i$ 、 $j$  间的连接权重进行调整:

$$\Delta\omega_{ij} = \eta(p_{ij}^{(+)} - p_{ij}^{(-)}) \quad (38.5-13)$$

式中  $\eta$ ——学习率;

$p_{ij}^{(+)}$ 、 $p_{ij}^{(-)}$ ——分别是  $i$  与  $j$  两个神经元在系统中处于  $\alpha$  状态和自由运转状态时实现连接的概率。

调整权重的原则是, 当  $p_{ij}^{(+)} > p_{ij}^{(-)}$  时, 则增加权重, 否则减少权重。权重调整的这种规则就是 Boltzmann 机的学习规则。

由于模拟退火过程要求高温, 使系统达到平衡状态, 而冷却退火过程又必须缓慢地进行, 否则易造成局部最小, 所以这种学习规则的主要缺点是学习速度很慢。

4) 竞争式学习 竞争式学习属于无教师学习方式。这种学习方式是利用不同层间的神经元发生兴奋性连接, 以及同一层内距离很近的神经元间发生同样的兴奋性连接, 而距离较远的神经元产生抑制性连接。这种在神经网络中的兴奋性或抑制性连接机制中引入了竞争机制的学习方式, 称为竞争式学习。它的本质特征在于神经网络中高层次的神经元对低层次神经元输入模式进行竞争式识别。

前向网络的竞争式学习规则, 是由 Rumelbar 和 Zipser 在 1985 年提出的。他们把前向网络结构设计成: 第一层为输入层, 而以后的每一层都增加许多不重叠的组块, 每一组块在特征识别中只有一个竞争优胜单元兴奋, 其余单元受到抑制。

设  $i$  为输入层某单元,  $j$  为获胜的特征识别单元, 则它们之间的连接权重变化为

$$\Delta\omega_{ij} = \eta[(C_{ik}/nk) - \omega_{ij}] \quad (38.5-14)$$

式中  $\eta$ ——学习率;

$C_{ik}$ ——外部层系列中  $i$  项刺激成分;

$nk$ ——刺激  $k$  激励输入单元的总数。

这种学习方式表明, 在竞争中输入单元间连接权重变化最大的优胜单元实现每一特征识别, 而失败的单元  $\Delta\omega_{ij}$  为零。

在竞争式学习机制的研究方面, 1987 年 Grossberg 提出将学习机制底-顶匹配和顶-底期望学习机制有机结合的原则, 进一步完善了他所建立的自适应共振网络模型 (ART)。ART 网络包括许多功能模块, 单个功能模块和单个层单元间按竞争式学习规则发生连接和变换。两层之间连接权重按照在竞争中优胜者取胜的原则进行调整。在多次顶-底与底-顶连接权重变化中优胜者取胜的平均值, 作为模式识别的分类模式标准。

采用竞争式学习机制的神经网络还有 Kohonen 提出的自组织特征映射网络等。

从上述介绍的几种学习规则不难看出, 要使人工神经网络具有学习能力, 就是使神经网络的知识结构变化, 这同把连接权重用什么方法变化是等价的。所以, 所谓神经网络的学习, 目前主要是指通过一定的学习算法对突触结合强度的调整, 使其达到具有记忆、识别、分类、信息处理和问题的优化求解等功能。当然, 使用 VLSI 技术, 在硬件上实现神经芯片, 可以极大提高神经网络的学习效率和进一步改善学习功能。

#### (4) 关于学习的其他问题

1) 学习和“随机波动” 在生物系统中, 经常出现随机波动。例如, 一个人即使发出同一音节, 第 1 回和第 2 回都存在微妙的差别。这样的差别, 在人类的行为中到处可见。在这里, 我们将其定义为“随机波动”。另一方面, 人类同计算机相比, 擅长于认识这些“随机波动”的现象。例如, 人类对于具有“随机波动”的声音和手写文字可以很容易地识别。

即使在学习中, 这种“随机波动”也起着重要的作用。例如, 在反复学习过程中, 我们期望有某种变化, 即“随机波动”。在生物进化过程中的学习、突然变异等实际上产生了某种“随机波动”, 形成了进化的契机。

2) 梦和学习 有的科学家提出: 人类的梦是在假睡眠 (Rapid Eye Movement, REM) 期间产生的, 而且假睡眠只有高级动物才会发生。

## 1.6 多层前向神经网络 (BP 网络)

### (1) 感知器

1957 年, 美国学者 Rosenblatt 提出了一种用于模式分类的神经网络模型, 被称为感知器 (perceptron)。它是由阈值元件组成且具有单层计算单元的神经网络, 其连接权重可变, 具有学习功能。

感知器信息处理的规则为

$$y(t) = f\left[\sum_{i=1}^n \omega_i(t)x_i - \theta\right] \quad (38.5-15)$$

式中  $y(t)$ —— $t$  时刻输出;

- $x_i$  —— 输入矢量的一个分量;
- $\omega_i(t)$  ——  $t$  时刻第  $i$  个输入的加权;
- $\theta$  —— 阈值;
- $f[\cdot]$  —— 阶跃函数。

感知器的学习规则如下:

$$\omega_i(t+1) = \omega_i(t) + \eta[d - y(t)]x_i \quad (38.5-16)$$

- 式中  $\eta$  —— 学习率 ( $0 < \eta < 1$ );
- $d$  —— 期望输出值 (又称教师信号);
- $y(t)$  —— 实际输出。

简单的 Perceptron 网络 (如图 38.5-17 所示) 为层状, 各层分别叫 S 单元 (sensory unit)、A 单元 (associative unit) 以及 R 单元 (Response unit)。根据以上学习规则, 其学习的流程如图 38.5-18 所示, 不断调整权重, 使得权重对于样本保持稳定不变, 学习过程即可结束。

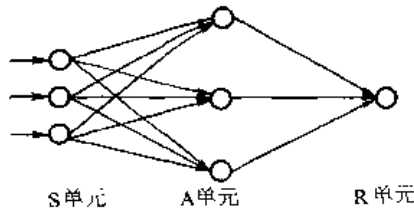


图 38.5-17 简单的 Perceptron 网络模型

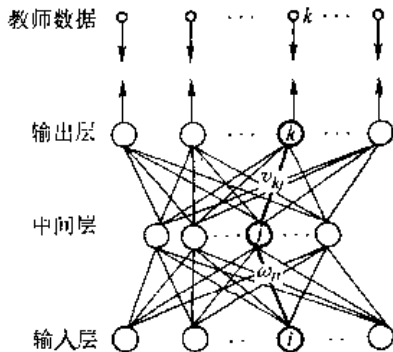


图 38.5-18 三层网络的结构

Rosenblatt 提出的感知器模型奠定了由信息处理、学习规则和作用函数三要素构成的基本模式。这种模型成为后来出现的几十种模型的重要基础。

应该指出, 上述的单层感知器能够解决一阶谓词逻辑问题, 如逻辑与、逻辑或问题, 但不能解决像异或问题的二阶谓词逻辑问题。感知器的学习算法保证收敛的条件, 要求函数是线性可分的 (指输入样本的函数类成员可分别位于直线分界线的两侧), 当输入函数不满足线性可分条件时, 上述算法受到了限制。为此, Minsky 等试图利用加入隐单元以扩大感知的功能, 但是由于缺乏有效的训练方法, 导致他们在 1969 年对感知器过分批评, 致使神经网络研究一度处于困境, 影响了研究工作的进展。含有隐单元的 Perceptron 网络

(见图 38.5-17) 为层状, 各层分别称 S 单元、A 单元和 R 单元。

(2) 前向多层网络的 BP 学习算法

前向网络是目前研究最多的网络形式之一, 如图 38.5-18 所示, 它包括输入层、隐层及输出层, 隐层可以为多层或一层。在隐层为一层的情况下也可把隐层叫作中间层。每层的神经元称为节点或单元。

BP (Back Propagation) 法是相对于由输入层、隐层和输出层构成的前向网络, 由 Rumelhart 等于 1986 年提出的有教师学习法。这个学习法同甘利 (1967 年)、Tsytkin (1966 年) 曾提出的概率下降法相同。

1) BP 网络误差反向传播学习算法的基本思想

- ① 向网络提供训练的例子, 包括输入单元的活性模式和期望的输出单元活性模式。
- ② 确定网络的实际输出与期望输出之间允许误差。

③ 改变网络中所有连接权重, 使网络产生的输出更接近于期望的输出, 直到满足确定的允许误差。

2) 误差反向传播算法的计算步骤 这里以图 38.5-18 所示的三层网络为例来说明误差反向传播算法的计算步骤, 见图 38.5-19。其中, 为了说明方便起见, 特引入一些记号。

① 初始化, 对所有连接权重赋予随机任意值, 并对阈值设定初值。

② 给定最初的学习数据。

③ 把学习数值  $\{I_i\}$  赋予输入层单元, 用从输入层到中间层之间的权重  $\{\omega_{ij}\}$  和中间层单元的阈值  $\theta_j$ , 求中间层单元  $j$  的输出  $U_j$ 。其中间层单元  $j$  的输入  $U_j$  和输出  $H_j$  由下式计算:

$$U_j = \sum_i \omega_{ij} \cdot I_i + \theta_j \quad (38.5-17)$$

$$H_j = f(U_j) \quad (38.5-18)$$

④ 用中间单元的输出  $\{H_j\}$  和从中间单元到输出单元的权重  $\{v_{kj}\}$ , 以及输出层单元  $k$  的阈值  $\gamma_k$ , 求输出层单元  $k$  的输入  $S_k$ 。其输出层单元  $k$  的输入  $S_k$  和输出  $O_k$  由下式计算:

$$S_k = \sum_j v_{kj} \cdot H_j + \gamma_k \quad (38.5-19)$$

$$O_k = f(S_k) \quad (38.5-20)$$

⑤ 由学习数据的教师信息  $T_k$  与输出层的输出  $O_k$  的差, 求出输出层单元  $k$  相对于中间层单元阈值的误差

$$\delta_k = (O_k - T_k) \cdot O_k \cdot (1 - O_k) \quad (38.5-21)$$

⑥ 用误差  $\delta_k$ 、从中间层向输出层的连接权重  $\{v_{kj}\}$  以及中间层的输出  $H_j$ , 求出相对于与中间层单元  $j$  连接的权重和中间层单元的阈值的误差

$$\sigma_j = \sum \delta_k \cdot v_k \cdot H_j \cdot (1 - H_j) \quad (38.5-22)$$

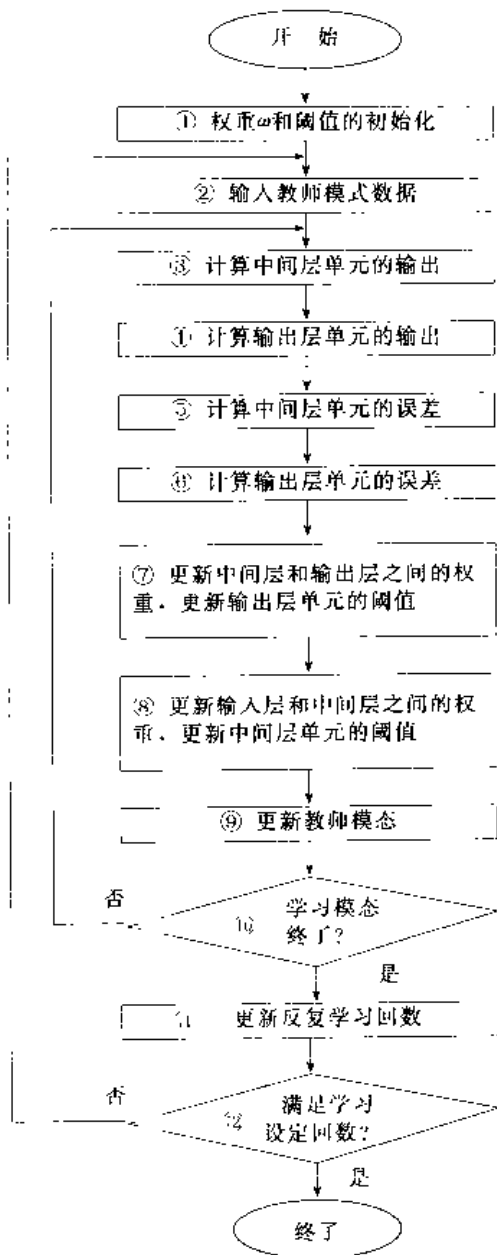


图 38.5-19 BP 学习法框图

⑦ 用输出层单元  $k$  的误差  $\delta_k$ 、中间层单元  $j$  的输出  $H_j$  以及系数  $\alpha$  的积的求和,修正从中间层单元  $j$  到输出层单元  $k$  连接权重  $v_{kj}$ 。同时用误差  $\delta_k$  和系数  $\beta$  的积的求和,修正输出层单元  $k$  的阈值  $\gamma_k$ :

$$v_{kj} = v_{kj} + \alpha \cdot \delta_k \cdot H_j \quad (38.5-23)$$

$$\gamma_k = \gamma_k + \beta \cdot \delta_k \quad (38.5-24)$$

⑧ 用中间层单元  $j$  的误差  $\sigma_j$ 、输入层单元  $i$  的输出  $I_i$  以及系数  $\alpha$  的积的求和,修正从输入层单元  $i$  到中间层单元  $j$  连接权重  $\omega_{ij}$ 。同时用误差  $\sigma_j$  和系数  $\beta$  的积的求和,修正中间层单元  $j$  的阈值  $\theta_j$ :

$$\omega_{ij} = \omega_{ij} + \alpha \cdot \sigma_j \cdot I_i \quad (38.5-25)$$

$$\theta_j = \theta_j + \beta \cdot \sigma_j \quad (38.5-26)$$

⑨ 把下一学习数据作为教师数据。

⑩ 学习数据终了的话,返回③。

⑪ 进行下一学习循环。

⑫ 假如学习次数不满足设定的学习次数,则返回②。

以上的③~⑥,是从输入层,经中间层到输出层的正向处理;⑦~⑧是从输出层,经中间层到输入层的反方向处理。因此,这种方法叫做 BP 算法。

总之,图 38.5-18 所示神经网络的输入输出关系,由图 38.5-20 所示的矩阵和矢量的关系可形象地表达。图 38.5-20 中有关记号的说明见表 38.5-2。

表 38.5-2 图 38.5-20 中记号的说明

记号	说 明
$I_i$	输入层单元 $i$ 的输出
$H_j$	中间层单元 $j$ 的输出
$O_k$	输出层单元 $k$ 的输出
$T_k$	相对于输出层单元 $k$ 的教师信息
$\omega_{ij}$	从输入层单元 $i$ 到中间层单元 $j$ 的连接权值
$v_{kj}$	从中间层单元 $j$ 到输出层单元 $k$ 的连接权值
$\theta_j$	中间层单元 $j$ 的阈值
$\gamma_k$	输出层单元 $k$ 的阈值

(3) BP 算法的问题以及改进算法

1) BP 算法的问题 BP 算法实质上是把一组样本输入输出问题转化为一个非线性优化问题,并通过梯度计算利用迭代运算求解权重问题的一种学习方法。已经证明,具有 Sigmoid 非线性函数的三阶神经网络可以以任意精度逼近任何连续函数。但是, BP 算法在实际应用中尚存在以下问题:

① 由于采用非线性梯度优化算法,易形成局部最小而得不到整体最优。

② 迭代算法次数甚多使得学习效率低,收敛速度很慢。

③ BP 网络是前向网络,无反馈连接,影响信息交换速度和效率。

④ 网络的输入节点、输出节点由问题而定,但隐节点的选取却根据经验,缺乏理论指导。

⑤ 在训练中学习新样本有遗忘旧样本的趋势,且要求表征每个样本的特征数目要相同。

⑥ 对每一应用,大多需要调整其学习参数。

2) BP 算法的改进

① MFBP 算法 BP 算法中网络参数每次调节的幅度,均以一个固定的因子  $\eta$  比例于网络误差函数或误差曲面,对这些参数的偏导数的幅度大小,是造成 BP 学习算法收敛很慢的一个重要原因。此外,网络参数的调节是沿着网络误差函数梯度下降的最快方向进行的,由于网络误差函数的 Hesse 矩阵出现严重的病态特征,致使这一梯度下降的最快方向极大地偏离指



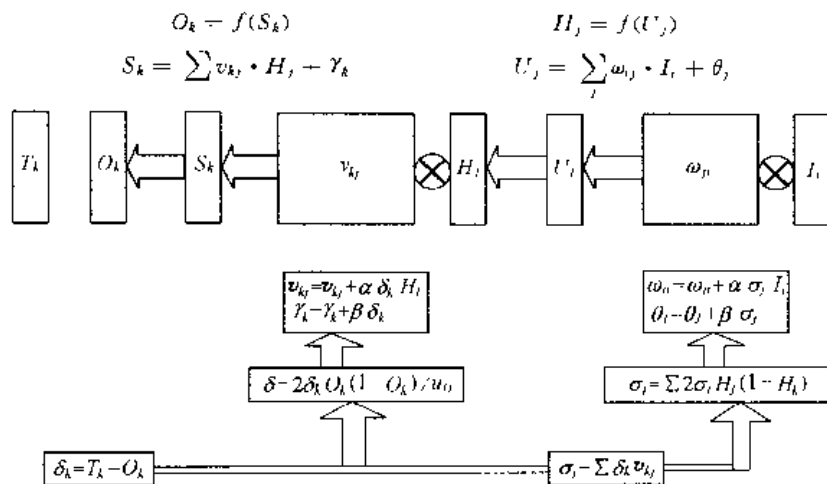


图 38.5-20 三层 BP 网络的计算流程

向误差曲面最小方向，从而急剧地加长了网络参数到达最小点位置的搜索路径，这又是造成 BP 学习算法收敛慢的另一个重要原因。其次，BP 学习过程中存在着某些“训练差的模式”，这些模式在学习过程中未对网络进行较好的训练，以致网络缺乏对这些模式的响应，从而退化网络的泛化特性，影响 BP 算法的学习效果。

克服这些缺陷的基本出发点是：网络中每个参数的调节应该有各自的学习率，且学习率在 network 整个学习过程中，可以根据网络误差曲面上的不同区域的曲率变化自适应调节。在误差曲面的某一区域处，若它对某一参数有较小的曲率，则在这一参数的连续几步调节中，误差函数对这一参数的偏导数一般具有相同的符号；若误差曲面相对于这一参数有较大的曲率，则在这一参数的连续几步调节中，误差函数对这一参数的偏导数符号一般将发生改变。因此，在这些参数的连续几步调节中，根据误差函数对网络参数的偏导数符号，可决定是否改变相应参数，来提高学习效率。为了强调于第 \$n\$ 步相邻偏导数符号的效应，采用指数加权平均方案。

经过分析推导可知，网络每层所对应的因子项的取值范围受到限制，是 Hesse 矩阵呈现病态的主要原因。为了减轻 Hesse 矩阵的病态特性，一个自然的想法就是将网络每层对应的病态因子

$$\prod_{m=1}^n O_{pkm} (1 - O_{pkm}) \quad n = 1, 2, \dots, L - 1$$

(L 为网络层数) (38.5-27)

取值范围扩至 \$[0, 1]\$ 区域。为此，网络中每个节点的输入输出关系修正为如下形式：

$$O_{pj} = 2 \left\{ 1 + \exp \left[ - 2 \left( \sum_i \omega_{ij} O_{pi} + \theta_j \right) \right] \right\}^{-1} - 1$$

(38.5-28)

其中：\$O\_{pj}\$ 为模式 \$p\$ 输至网络节点 \$j\$ 的输出，同时网络的节点误差信号 \$\delta\_{pj}\$ 相应变为

$$\delta_{pj} = (t_{pj} - O_{pj})(1 + O_{pj})(1 - O_{pj}) \quad (j \text{ 为输出节点})$$

(38.5-29)

$$\delta_{pj} = \sum_k \delta_{pk} (1 + O_{pj})(1 - O_{pj}) \quad (j \text{ 为隐节点})$$

(38.5-30)

基于以上讨论，快速算法 FBP 算法由式 (38.5-29)、式 (38.5-30) 及以下公式描述：

$$\omega_{pj}(n+1) = \omega_{pj}(n) + \eta_{pj}(n) \sum_p \delta_{pj}(n) O_{pj}(n)$$

(38.5-31)

其中：

$$\eta_{pj}(n-1) = \begin{cases} \eta_{pj}(n) \cdot \alpha \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \Delta_1(n-1) > 0 \\ \eta_{pj}(n) \cdot \beta \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \Delta_1(n-1) < 0 \\ \eta_{pj}(n) \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \Delta_1(n-1) = 0 \end{cases}$$

$$\Delta_1 = \gamma \cdot \Delta_1(n-1) + (1 - \gamma) \left[ \sum_p \delta_{pj}(n) O_{pj}(n) \right]$$

$$\Delta_1(0) = \sum_p \delta_{pj}(n) O_{pj}(n)$$

$$\theta_j(n+1) = \theta_j(n) + \eta_j(n) \sum_p \delta_{pj}(n) \quad (38.5-32)$$

其中：

$$\eta_j(n-1) = \begin{cases} \eta_j(n) \cdot \alpha \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \cdot \Delta_2(n-1) > 0 \\ \eta_j(n) \cdot \beta \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \cdot \Delta_2(n-1) < 0 \\ \eta_j(n) \cdot \left[ \sum_p \delta_{pj}(n) O_{pj} \right] \cdot \Delta_2(n-1) = 0 \end{cases}$$

$$\Delta_2(n) = \gamma \cdot \Delta_2(n-1) + (1 - \gamma) \left[ \sum_p \delta_{pj}(n) \right]$$

$$\Delta_2(0) = \sum_p \delta_{pj}(n)$$

其中：\$\alpha > 1, 0 < \beta < 1, 0 < \gamma < 21\$，均为一选定的常数因子。\$\omega\_{pj}(0)\$ 与 \$\theta\_j(0)\$ 预先初始化为 \$[-1, 1]\$ 内均匀分

布的随机数,  $\eta_p(0)$  与  $\eta_i(0)$  为预先给定的一小的正数。

采用上述 FBP 算法对收敛速度有大幅度的提高, 进一步分析 FBP 算法, 可以看出训练网络的学习时间和其推广特性与训练模式在学习过程中的行为有密切的关系, 于是在 FBP 算法扩大的基础上, 采用动态训练集技术可改变网络的推广特性, 这样就构成了 MFBP 算法:

(a) 用  $P$  中所有元素按 FBP 算法对网络进行一次调节后, 可得集合

$$T \Delta \{p \in P; E_p / \#P\}$$

其中:  $\#P$  表示集合  $P$  中元素的个数;  $E_p$  为训练集中每个样本模式  $p$  所产生的网络误差。

(b) 建立动态训练集 DST。设其中至少有  $d$  个元素, 它由集合  $P$  中所有元素和非主导训练集合  $T$  中所有元素或其复制元素构成。如果  $P$  中元素个数与  $T$  中元素个数之和大于或等于  $d$ , 则 DST 即由  $P$  中所有元素和  $T$  中所有元素构成; 如果  $P$  中元素个数与  $T$  中元素个数之和小于  $d$ , 则 DST 由  $T$  中所有元素、 $P$  中所有元素和  $T$  中某些元素的复制元素构成, 其中  $T$  中某些元素复制多少应视 DST 中元素个数是否达到  $d$  而定。

(c) 用 DST 集合对网络参数按 FBP 进行调节, 检查 FBP 算法收敛条件是否满足, 如果满足, 则结束训练, 否则转向①。

(d) 更新 DST, 在  $P$  中寻找具有最大误差值  $E_p$  的元素  $p$ , 在  $T$  中寻找具有最小误差值  $E_t$  的元素  $t$ , 如果  $t=p$ , 则在 DST 中增加一个元素  $p$  的复印元素, 使 DST 扩大; 如果  $t \neq p$ , 则用  $p$  取代  $T$  中的元素  $t$ 。

(e) 转向③

(f) 训练结束。

MFBP 算法比 BP 算法具有更快的收敛速度。仿真结果表明, MFBP 算法的迭代次数为 BP 算法的  $1/9 \sim 1/7$ , 且具有更好的推广特性。

### ② MBP 算法

MBP 算法通过改变作用函数  $f(x)$  的值域, 即加入一个增益因子  $C$  以改变作用函数陡度, 在训练过程中, 增益因子  $C$  随权值  $\omega$  和阈值  $\theta$  一起发生变化, 以达到改善 BP 算法的收敛特性, 加快收敛速度的目的。利用牛顿最速下降法可得到 BP 算法如下:

$$\delta_{jk} = f(Net_{jk}) \sum_m \delta_{mk} \omega_{mj} \quad (38.5-33)$$

$$\frac{\partial E}{\partial \omega_{ij}} = \delta_{ij} O_{ik} \quad (38.5-34)$$

$$\omega_{ij}(t+1) = \omega_{ij} - \mu \frac{\partial E}{\partial \omega_{ij}} \quad (38.5-35)$$

其中:

$$Net_{jk} = \sum_i \omega_{ij} O_{ik} \quad (38.5-36)$$

$$O_{ik} = f(Net_{jk}) \quad (38.5-37)$$

$$\delta_{ik} = -\frac{\partial E_k}{\partial Net_{ik}} \quad (38.5-38)$$

$$\frac{\partial E}{\partial \omega_{ij}} = \sum_{k=1}^n \frac{\partial E_k}{\partial \omega_{ij}} \quad (38.5-39)$$

$\mu$  为学习因子,  $\mu > 0$ 。

利用式 (38.5-33) ~ 式 (38.5-39) 可推得网络权值和阈值的学习算法为

$$\Delta \omega_{ij}(t+1) = \eta \delta_{jk} O_{ij} + a \Delta \omega_{ij}(t) \quad (38.5-40)$$

$$\Delta \theta_j(t+1) = \eta \delta_{jk} - a \Delta \theta_j(t) \quad (38.5-41)$$

式中  $\eta$  —— 学习步长;

$a$  —— 权值记忆因子。

分析式 (38.5-40)、式 (38.5-41) 可知, 影响 BP 算法的因素很多, 归纳起来主要包括学习步长  $\eta$ 、权值记忆因子  $a$ 、网络结构及节点作用函数等。这些参数主要根据经验选取, 具有一定的局限性, MBP 算法主要基于这些参数的调整, 对 BP 算法进行改进。

首先, 将节点作用函数修改为

$$f(x) = -0.5 + \frac{1}{1 + e^{-x}} \quad (38.5-42)$$

此时, 函数的值域由  $(0, 1)$  变为  $(-0.5, 0.5)$ 。由此可对零输入样本进行学习训练, 能克服 BP 算法中零输入样本时, 相关的权值和阈值均不改变计算的有效问题, 从而加快收敛速度。

其次, 适当增加作用函数的陡度, 以利于改善算法的收敛性。为此在节点净输入  $Net_i$  前加一个常数因子  $C^m$ , 这样输出函数  $y_i^m = f(Net_i^m)$  变为

$$y_i^m = f(C^m Net_i^m) \quad (38.5-43)$$

最后, 对学习步长采用变步长策略, 可根据收敛性加以调整。即在收敛过程中, 本次误差大于上次误差, 则这次迭代无效, 恢复迭代前的步长, 减少步长增加的幅度重新迭代; 反之, 本次迭代有效, 增大学习步长。

下面给出 MBP 算法的具体公式及简单的推导过程。

对于一个多层网络, 假设  $y_i^m$  是第  $m$  层的第  $i$  个节点的输出,  $\omega_{ij}^m$  是第  $m$  层的第  $i$  各节点到第  $m-1$  层的第  $j$  个节点的连接权值, 第  $m$  层的第  $i$  个节点的净输入  $Net_i^m = \omega_{ij}^m = \sum_k \omega_{ik}^m y_k^{m-1}$ , 这时,  $Net_i^m = [Net_1^m, Net_2^m, \dots, Net_n^m]$  是第  $m$  层的净输入的列矢量, 节点作用函数为式 (38.5-42)、式 (38.5-43)。

假如对于一个输入模式  $x_0$ , 期望的输出为  $y = [y_1, y_2, \dots, y_n]^T$ , 而实际的输出为  $\hat{y}$  对于二次型误差函数

$$E = \frac{1}{2} \sum_{k=1}^n (y_k - \hat{y}_k) \quad (38.5-44)$$

将  $E$  对  $\omega_i^m$  取偏微分得

$$\frac{\partial E}{\partial \omega_i^m} = \frac{\partial E}{\partial Net^m} \cdot \frac{\partial Net^m}{\partial y_i^m} \cdot \frac{\partial y_i^m}{\partial Net_i^m} \cdot \frac{\partial Net_i^m}{\partial \omega_i^m}$$

$$= [-\delta_1^{m-1} \quad \delta_2^{m-1} \quad \dots \quad -\delta_n^{m-1}]$$

$$\begin{bmatrix} \omega_{i1}^m \\ \omega_{i2}^m \\ \vdots \\ \omega_{in}^m \end{bmatrix} f'(C^m Net_i^m) x_i^{m-1}$$

(38.5-45)

其中:

$$\delta_i^m = -\partial E / \partial Net_i^m \quad (38.5-46)$$

当  $i$  为隐节点时

$$\delta_i^m = \left( \sum_k \delta_k^{m-1} \omega_{ki}^{m-1} \right) f'(C_i^m Net_i^m) C_i^m$$

(38.5-47)

当  $i$  为输出节点时

$$\delta_i^m = (t_{ip} - O_{ip}) f'(C_i^m Net_i^m) C_i^m \quad (38.5-48)$$

根据式 (38.5-46)、式 (38.5-41) 可得

$$\Delta \omega_{ij}^m = \rho_\omega \delta_i^m y_j^{m-1} \quad (38.5-49)$$

其中:  $\rho_\omega$  为权的学习步长。

同理, 计算误差对增益常数  $C$  的梯度

$$\frac{\partial E}{\partial C_i^m} = \left( \sum_k \delta_k^{m-1} \omega_{ki}^{m-1} \omega_{ki}^{m-1} \right) f'(C_i^m Net_i^m) Net_i^m$$

(38.5-50)

$$\Delta C_i^m = \rho_c \delta_i^m \cdot Net_i^m / C_i^m \quad (38.5-51)$$

其中:  $\rho_c$  为增益的学习步长, 只要将  $\rho_c$  取为 0,  $C_i^m$  初值取为 1, 增益  $C_i^m$  不再起作用。

MBP 算法经过用于系统辨识仿真表明, 与 BP 算法相比, 在选用相同的学习训练步数条件下, MBP 算法的精度远远高于 BP 算法的精度 (至少高出三个数量级)。对于相同精度的要求, 则 MBP 算法可在很少步数内达到精度要求, 收敛特性、收敛速度及系统的泛化特性都得到了改善。

### ③ 前向网络的自构形学习算法

在前向网络的拓扑结构中, 输入节点与输出节点数是由问题本身决定的。而隐节点数的选取则相对困难得多。隐节点数少了, 学习过程不收敛; 隐节点数多了, 存在节点冗余, 网络性能下降。为了找到合适的隐节点数, 最好的办法是网络在学习过程中, 根据环境的要求, 自组织和自学习自己的结构, 这种网络学习的方法称为自构形学习算法。这种学习过程分为预估和自构形两个阶段。预估就是根据问题的大小及复杂程度, 设定一个隐节点数很大的前向网络结构。在自构形阶段, 网络根据学习情况合并无用的冗余节点, 最后得到

一个合适的自适应网络。

设  $O_{ip}$  是隐节点  $i$  在学习第  $p$  个样本时的输出,  $\bar{O}_i$  和  $\bar{O}_j$  是隐节点  $i$  和  $j$  在学习完  $n$  个样本后的平均输出,  $n$  为训练样本总数, 则

$$\bar{O}_i = \frac{1}{n} \sum_{p=1}^n O_{ip} \quad (38.5-52)$$

$$\bar{O}_j = \frac{1}{n} \sum_{p=1}^n O_{jp} \quad (38.5-53)$$

为了衡量隐节点的工作情况, 给出同层隐节点间的相关系数及样本分散度的定义如下:

同层隐节点  $i$  和  $j$  的相关系数

$$\gamma_{ij} = \frac{\left( \frac{1}{n} \sum_{p=1}^n O_{ip} O_{jp} - \bar{O}_i \bar{O}_j \right)}{\left( \frac{1}{n} \sum_{p=1}^n O_{ip}^2 - \bar{O}_i^2 \right) \left( \frac{1}{n} \sum_{p=1}^n O_{jp}^2 - \bar{O}_j^2 \right)}$$

(38.5-54)

$\gamma_{ij}$  表明隐节点  $i$  和  $j$  输出的相关程度,  $\gamma_{ij}$  过大, 说明节点  $i$  和  $j$  功能重复, 需要压缩合并。

样本分散度为

$$S_i = \frac{1}{n} \sum_{p=1}^n O_{ip}^2 - \bar{O}_i^2 \quad (38.5-55)$$

$S_i$  过小, 表明隐节点  $i$  的输出值变化很小, 它对网络的训练没有起作用, 性能类同于阈值。

基于以上定义, 给出隐节点动态合并和删减规则如下:

① 合并规则: 若  $|\gamma_{ij}| \geq c_1$ , 且  $S_i, S_j \geq c_2$ , 则同层隐节点  $i$  和  $j$  可以合而为一。这里  $c_1$  和  $c_2$  为规定的下限值。一般  $c_1$  取 0.8 ~ 0.9,  $c_2$  取 0.001 ~ 0.010。令  $O_j \approx aO_i + b$ , 则

$$a = \frac{\left( \frac{1}{n} \sum_{p=1}^n O_{ip} O_{jp} - \bar{O}_i \bar{O}_j \right)}{\left( \frac{1}{n} \sum_{p=1}^n O_{ip}^2 - \bar{O}_i^2 \right)}$$

$$b = O_j - aO_i$$

如图 38.5-21 所示, 输出节点  $k$  的输入

$$Netin_k = \omega_{ki} O_i + \omega_{kj} O_j + \omega_{kb} \cdot 1 + \sum_{l \neq i, j} \omega_{kl} O_l$$

$$= (\omega_{ki} + a\omega_{kj}) O_i + (\omega_{kb} + b\omega_{kj}) \cdot 1 + \sum_{l \neq i, j} \omega_{kl} O_l$$

从而得到合并算法:  $\omega_{ki} \rightarrow \omega_{ki} + a\omega_{kj}$ ;  $\omega_{kb} \rightarrow \omega_{kb} + b\omega_{kj}$ 。

② 删减规则: 若  $S_i < c_2$ , 则节点  $i$  可以删除。令  $O_i \approx \bar{O}_i$ , 则输出节点  $k$  的输入

$$Netin_k = \omega_{ki} O_i + \omega_{kb} \cdot 1 + \sum_{l \neq i, j} \omega_{kl} O_l$$

$$= (\omega_{kb} + \omega_{ki} \bar{O}_i) \cdot 1 + \sum_{l \neq i, j} \omega_{kl} O_l$$

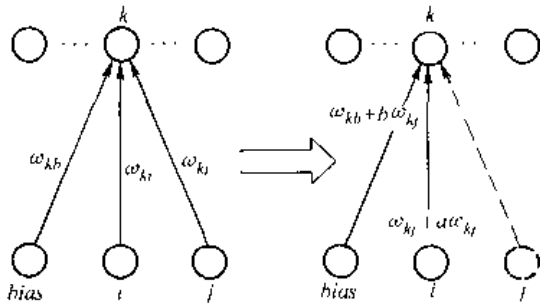


图 38.5-21 隐节点  $i$  和  $j$  的合并

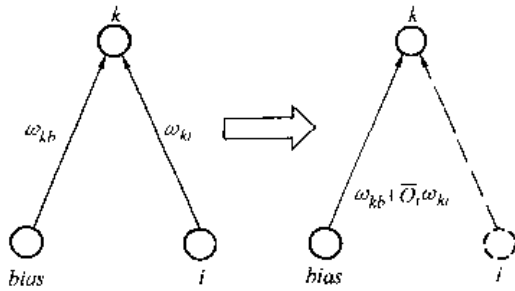


图 38.5-22 隐节点  $i$  的删减

所以删减算法为： $\omega_{kb} \rightarrow \omega_{kb} + \theta_i \omega_{ki}$ ，如图 38.5-22 所示，实际上是节点和阈值节点合并了。

根据上述的定义和规则可得到前向网络的自构形学习算法。采用三层前向网络应用自构形算法，对机械加工工件特征识别实验结果表明，这种算法同 BP 算法相比，不仅识别准确率高，而且收敛时间缩短约 1/7~1/5。

### 1.7 典型反馈网络——Hopfield 网络

前面讨论的前向网络是单向连接没有反馈的静态网络，从控制系统的观点看，它缺乏系统动态性能。美国物理学家 Hopfield 对神经网络的动态性能进行深入研究，在 1982 年首先提出了一种由非线性元件构成的单层反馈网络系统，称为 Hopfield 网络。从控制系统的观点看，Hopfield 网络是一个非线性动力学系统，由于非线性系统本身的复杂性，且涉及到随机性、稳定性、吸引子以至于混沌现象等问题，所以使得研究反馈网络要比前向网络复杂得多。

#### (1) Hopfield 网络的物理学模型

1982 年美国物理学家 J. J. Hopfield 发表了名为“神经网络与基于自然发生集合计算能力的物理系统”的论文。在论文中提出了一种相互连接型网络模型，并指出其模型同物理学中说明物质的磁性的磁极模型非常相似。物理学中的磁极，以“向上”和“向下”两种状态为磁性的基本单位，由多数磁极间的相互作用而产生磁力。例如，永久磁铁的磁力，就是由许多磁极朝向相同的相互作用而产生。

假如把磁极模型的磁极置换为神经元，把磁极间

相互作用置换为神经元间的连接，则得到的神经网络即为 Hopfield 的基本网络（见图 38.5-23）。磁极的向上和向下状态，正好与神经元的兴奋和抑制相对应。同时，为使 Hopfield 网络与物理学中的磁极模型有关概念相对应，引入了神经网络的能量的思想。使用能量这一量度，以考察神经网络的动作以及其全局的信息处理能力。

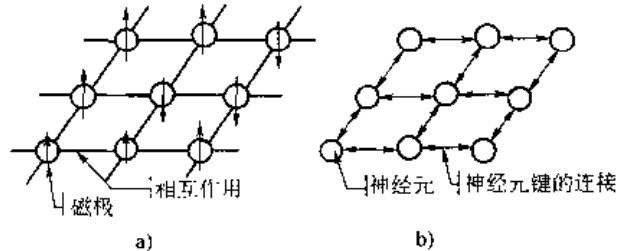


图 38.5-23 磁极模型和神经网络模型的对比

a) 磁极 b) 神经网络

#### (2) Hopfield 网络模型

Hopfield 网络模型的拓扑结构可看作全连接加权无向图，它是一种网状网络，可分为离散和连续两种类型。离散网络的节点仅取 +1 和 -1（或 0 和 1）两个值，而连续网络取 0 和 1 之间任一实数。其信息流向是双向的，即在网络中，单元自己的输出作为其他单元的输入；其他单元的输出又作为自己的输入，从而实现反馈。图 38.5-24 给出 Hopfield 网络的一种结果形式。

1) Hopfield 网络模型基本特征的两个重要约束条件

① 对称的相互连接性 从神经元  $i$  到  $j$  间的连接权值  $\omega_{ij}$  与从神经元  $j$  到  $i$  间的连接权值  $\omega_{ji}$  相等即为对称的相互连接性，如图 38.5-25a 所示。因为在实际的生物神经网络中，神经元间的突触连接强度完全是非对称的，如图 38.5-25b 所示，所以其对称性作为生物学的模型也许难以被人们接受。但是，从 Hopfield 的论文名可知，作为新的计算机（信息处理）模型的神经网络，其神经元间对称相互连接的约束条件的引入，大概不会有大的问题。其次，这种对称性同 Hopfield 参照的物理学磁极模型之间具有类似性。

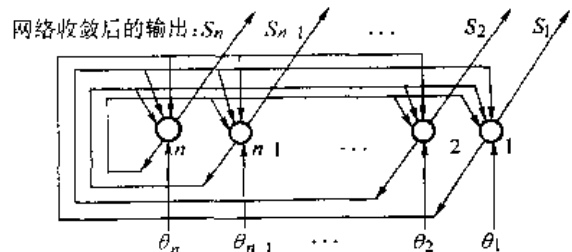


图 38.5-24 Hopfield 网络

② 非同期的神经元动作 神经网络中的各神经元以非同期的形式进行动作即为非同期的神经元动

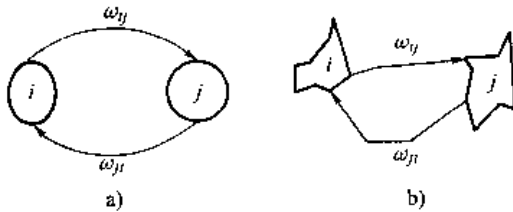


图 38.5.25 对称的相互连接

作。以前的计算机,其所有的理论单元都是以时钟同期动作,进行具有系统意义的某些处理。Hopfield 网络中,各神经元以自己的动态改变自己的状态(输出)的形式来实现系统信息处理。这种各神经元完全独立地动作,作为并行计算机的模型是非常具有魅力的。

在以上两个约束条件下,设此网络含有  $n$  个神经元,在时刻  $t$ ,神经元  $i$  由其他  $n-1$  个神经元得到的输入信息的总和为

$$I_i(t) = \sum_{j \neq i} \omega_{ij} S_j(t) + \theta_i \quad (38.5-56)$$

其中:  $\omega_{ij}$  为神经元  $i$  与  $j$  间的连接权值;  $\theta_i$  为其阈值。在时刻  $t+1$ ,神经元  $i$  的状态(输出)值  $S_i(t+1)$  取 0 或 1,各神经元按下列规则随机地、异步地改变状态:

$$S_i(t+1) = \begin{cases} 1 & I_i \geq 0 \\ 0 & I_i < 0 \end{cases} \quad (38.5-57)$$

## 2) Hopfield 网络的时间发展规则

① 随机地从网络中选出 1 个神经元。

② 被选中的神经元  $i$  ( $1 \leq i \leq n$ ),其输入的总和

为

$$I_i(t) = \sum_{j \neq i} \omega_{ij} S_j(t) + \theta_i$$

③ 基于  $I_i(t)$  的值对神经元  $i$  的输出值进行更新:

$$S_i(t+1) = \begin{cases} 1 & (I_i(t) \geq 0) \\ 0 & (I_i(t) < 0) \end{cases}$$

④ 神经元  $i$  以外的神经元  $j$  的输出不让其发生变化:

$$S_j(t-1) = S_j(t)$$

⑤ 返回①。

## (3) 网络的能量

Hopfield 根据系统动力学和统计学原理,在网络系统中引入能量函数的概念,给出了网络系统稳定性问题定理(Hopfield 定理)。

**定理** 设  $(\omega, S)$  是神经网络,若  $\omega_{ij} = \omega_{ji}$ ,且  $\omega_{ij} > 0$ ,同时各神经元是随机异步地改变状态,则此网络一定能收敛到稳定的状态。

**证** 在网络中引入能量函数

$$E(t) = -\frac{1}{2} \sum_{i,j} \omega_{ij} S_i(t) S_j(t) - \sum_i \theta_i S_i(t) \quad (38.5-58)$$

能量函数  $E(t)$  随状态  $S_k(t)$  的变化为

$$\frac{\Delta E(t)}{\Delta S_k(t)} = -\frac{1}{2} \sum_{j=1}^n \omega_{kj} S_j(t) - \frac{1}{2} \sum_{j=1}^n \omega_{jk} S_j(t) + \theta_k \quad (38.5-59)$$

当  $\omega_{ik} = \omega_{ki}$  时,则得

$$\Delta E_k(t) = -\left[ \sum_{j=1}^n \omega_{kj} S_j(t) + \theta_k \right] \Delta S_k(t) \quad (38.5-60)$$

若令  $S_k(t) = \sum_{j=1}^n \omega_{kj} S_j(t) + \theta_k$ ,则由式(38.5-56)

可知,  $S_k(t)$  与  $\Delta S_k(t)$  同号,再从式(38.5-60)可以看出,  $S_k(t)$  与  $\Delta S_k(t)$  之积大于零,故  $\Delta E_k(t) < 0$ 。这表明网络系统总是朝着能量减少的方向变化,最终进入稳定状态。

Hopfield 网络模型的基本原理是:只要由神经元兴奋的算法和连接权系数所决定的神经网络的状态,在适当给定的兴奋模式下尚未达到稳定状态,那么该状态就会一直变化下去,直到预先定义的一个必定减小的能量函数达到极小值时,状态才达到稳定而不再变化。

如果更直观地赋予这种反馈网络系统的动态特性,则整个网络系统中神经元的的信息处理过程就好像与在满员的电车中拥挤的乘客最初以不自然的姿势相互拥挤着,后来逐渐地以安定的姿势稳定下来的过程一样,犹如神经元由混沌状态转移到稳定状态。

对于连续的 Hopfield 网络,通过引入能量函数,可以类同于离散模型的情况加以研究,它同样朝着能量减少的方向运行,最终达到一个稳定的状态。

## (4) Hopfield 网络的联想记忆功能

如前所述, Hopfield 网络是一个非线性动力学系统。引入能量函数后并从系统能量的观点看,该网络系统在一定的条件下,总是朝着系统能量减少的方向变化,并最终达到能量函数的极小值而不再变化。如果把这个极小值所对应的模式作为记忆模式,那么以后当给这个网络系统一个适当的激励时,它就能成为想起已记忆模式的一种联想记忆模式,即 Hopfield 网络具有联想记忆功能。

Hopfield 网络联想记忆过程,从动力学的角度就是非线性动力学系统朝着某个稳定状态运行的过程,这需要调整连接权值使得所要记忆的样本作为系统的吸引子,即能量函数的局部最小点。这一过程可分为学习和联想两个阶段。

在给定样本的条件下,按照 Hebb 学习规则,调整连接权值,使得存储的样本成为动力学的吸引子,这个过程就是学习阶段。而联想是指在已调整好权值不变的情况下,给出部分不全或受了干扰的信息,按照动力

学规则改变神经元的状态,使系统最终变到动力学的吸引子。也即指收敛于某一点,或周期性迭代(极限环),或处于混沌状态。

Hopfield 网络模型的动力学规则是指:若网络节点在  $S(0)$  初始状态下,经过  $t$  步运行后将按下述规则达到  $S(t+1)$  状态,即

$$S_i(t+1) = \text{sgn} \left[ \sum_{j=1}^n \omega_{ij} S_j(t) + \theta_i \right] \quad (38.5-61)$$

其中:  $\text{sgn}$  为符号函数。

实现 Hopfield 网络联想记忆的关键是,网络到达记忆样本能量函数的极小点时,决定网络的神经元间连接权值  $\omega_{ij}$  和阈值  $\theta_i$  等参数。以下将 Hopfield 网络用于联想记忆的学习算法规则作一整理。

1) 按照 Hebb 规则设置权值

$$\omega_{ij} = \begin{cases} \sum_{m=1}^n x_i^m x_j^m & i \neq j \\ 0 & i = j \end{cases} \quad i, j = 1, 2, \dots, n \quad (38.5-62)$$

其中:  $\omega_{ij}$  是节点  $i$  到节点  $j$  的连接权值;  $x_i^m$  表示样本集合  $m$  中的第  $i$  个元素;  $x_i \in [-1, +1]$ 。

2) 对未知样本初始化

$$S_i(0) = x_i \quad i = 1, 2, \dots, n \quad (38.5-63)$$

其中:  $S_i(t)$  是时刻节点  $i$  的输出;  $x_i$  是未知样本的第  $i$  个元素。

3) 迭代计算

$$S_j(t+1) = \text{sgn} \left[ \sum_{i=1}^n \omega_{ij} S_i(t) + \theta_j \right] \quad \theta_j > 0, j = 1, 2, \dots, n \quad (38.5-64)$$

直至节点输出状态不改变时,迭代结束。此时节点的输出状态即为未知输入最佳匹配的样本。

4) 返回 2) 继续迭代。

依据上述算法的联想记忆功能,可用于模式识别,但样本较多且彼此间相近时,容易引起混淆。因此,这种网络要求存储的信息模式必须是两两正交。此外,网络要求连接权中满足对称条件等。

(5) Hopfield 网络的优化计算功能

Hopfield 网络理论的核心思想为:网络从高能状态转移到最小能量状态,则达到收敛,获得稳定的解,完成网络的功能。Hopfield 网络所构成的动力学系统与固体物理学模型自旋玻璃相似,可用二次能量函数来描述系统的状态,系统从高能状态到低能的稳定状态的变化过程,相似于满足约束问题的搜索最优化解的过程。因此, Hopfield 网络可用于优化问题的计算。

Hopfield 网络用于优化问题的计算与用于联想记忆的计算过程是对偶的。在解决优化问题时,权矩阵  $\omega$

已知,目的是求取最大能量  $E$  的稳定状态。为此,必须将优化的问题映射到网络的响应于优化问题可能解的特定组态上,再构造优化问题的能量函数,它应和优化问题中的二次型代价函数成正比。

通过将能量函数和代价函数相比较,求出能量函数中权值和偏值,并以此去调整响应的反馈权值和偏值,进行迭代计算,直到系统收敛到稳定状态为止。最后将所得到的稳定状态变换为实际优化问题的解。

Hopfield 应用这种网络优化计算功能,成功地解决了著名的旅行商 TSP 问题。TSP 问题是给定  $N$  个城市,从某一城市出发走遍所有城市但不准重复,然后再回到原出发地其间的路径必须最短。此问题易于表达,但难于求解。用普通搜索法极其费时,而采用 Hopfield 网络在极短时间内找到虽不是最短,但却是接近最短的最优近似解,充分显示出这一方法的巨大潜力。

求解 TSP 问题,其关键在于对所求问题找出用合适的数学表达的能量函数的形式,网络的行为须用能量函数  $E$  来描述,它的最小值响应于最佳路径。

Hopfield 网络理论不仅奠定了用动力学和统计力学理论研究反馈网络的基础,而且连续的 Hopfield 网络模型可直接与电子模拟线路相对偶,易于通过 VLSI 技术实现,这样就为研制神经计算机开辟了道路。

## 1.8 基于概率学习的 Boltzmann 机模型

1985 年加拿大多伦多大学教授 G. E. Hinton 等人基于统计物理学和 Boltzmann 提出的概率分布模拟退火过程,提出了 Boltzmann 机的学习算法。这个模型是把 Hopfield 网络的动作规则,以概率动作的形式扩张其规则的网络模型。此扩张,使 Hopfield 网络中难以确定的神经元间的连接权值及网络参数等,由学习可以予以确定。其次,网络的动作也不是收敛于能量函数的极小点,而可能收敛于能量函数的最小点。

(1) Boltzmann 机模型

Boltzmann 机是一个相互连接的神经网络模型,其特点是隐节点间具有相互结合的关系。每个都根据自己的能量差  $\Delta E$  随机地改变自己的或为 1 或为 0 的状态。

前述的 Hopfield 网络的时间发展规则可以归结如下:

$$I_i(t) = \sum_{j \neq i} \omega_{ij} S_j(t) + \theta_i$$

$$S_i(t+1) = 1(I_i(t))$$

其中:  $S_i(t)$  是包括输入层、隐层和输出层中所有神经元的状态;函数  $1(\cdot)$  是阶跃函数(也称为阈值函数),可表示为

$$l(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

Boltzmann 机在进行  $\omega_j, I_i(t), S_i(t-1)$  的计算中,以概率判定的形式取代阶跃函数。神经元  $i$  在下一时刻的输出值  $S_i(t+1)$ ,根据  $I_i(t)$  的值计算式(38.5-65)概率  $P$ ,确定其是否为 1。

$$P(S_i(t+1) = 1) = f(I_i(t)/T) \quad (38.5-65)$$

其中,  $f(x)$  为 Sigmoid 函数,其表达式为

$$f(x) = \frac{1}{2} \left( 1 + \tanh \left( \frac{x}{2} \right) \right) = \frac{1}{1 + \exp(-x)} \quad (38.5-66)$$

其中,  $T$  为网络的温度。随着  $T$  的变化,函数  $f(x/T)$  的变化形式如图 38.5-26 所示。定性地说,  $T$  的值越大(温度越高),  $f(x/T)$  函数相对于  $x$  越不敏感,  $T \rightarrow \infty$  时,  $f(x/T) = 0.5$ ;  $T$  的值小(温度低),  $f(x/T)$  函数相对于  $x$  的值出现正负,表现出敏感性。  $T \rightarrow 0$  时,  $f(x/T)$  为阶跃函数。这时,  $I_i(t)$  的值为正,用概率 1.0,  $S_i(t-1) = 1$ ;  $I_i(t)$  的值为负,用概率 1.0,  $S_i(t+1) = 0$ 。

由此,我们明确了在与 Hopfield 网络的动作规则相同且  $T=0$  的前提下, Boltzmann 机与 Hopfield 网络是一致的。以下,我们来讨论用式(38.5-65)所给的概率条件,向  $S_i(t+1) = 1$  状态迁移时网络能量的变化。在网络状态迁移前后,因为输出的变化仅局限于神经元  $i$ ,由式(38.5-58)状态迁移前后的能量变化量为

$$\Delta E = E(t+1) - E(t) = - (1 - S_i(t)) \cdot I_i(t) \quad (38.5-67)$$

由上式可知,在  $I_i(t)$  为正的情况下,随着状态迁移,其能量减少(或不变化)。由图 38.5-26 可知,  $I_i(t)$  为正的情况下,发生状态迁移的概率也大。相反地,在  $I_i(t)$  为负的情况下,随着状态迁移,其能量增加(或不变化)。这种迁移在 Hopfield 网络中虽是禁止的,但在 Boltzmann 机中允许以小概率出现(图 38.5-26 中横轴的左半部分)。正是由此原因, Boltzmann 机的状态迁移函数不停留在极小点,而可能收敛于最小点。当系统达到平衡时,能量函数达到最小值。可以证明 Boltzmann 机是收敛的。将 Boltzmann 机的时间发展规则整理如下:

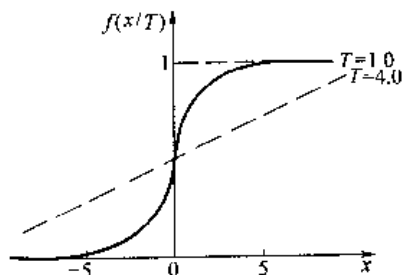


图 38.5-26 Sigmoid 函数

1) 随机地从网络中选出 1 个神经元。

2) 被选中的神经元  $i(1 \leq i \leq n)$ ,其输入的总和为

$$I_i(t) = \sum_{j \neq i} \omega_{ij} S_j(t) + \theta_i$$

3) 以式(38.5-65)的概率  $P$ ,把神经元  $i$  输出值  $S_i(t+1)$  设定为 1。

4) 神经元  $i$  以外的神经元  $j$  的输出不让其发生变化,

$$S_j(t+1) = S_j(t)$$

5) 返回 1)。

(2) 模拟退火

模拟退火的基本思想源于统计力学,统计力学是研究一个多自由度的系统在某温度下达到热平衡时的行为特性。金属在高温熔化时,所有原子都处于高能自由运动状态,随着温度的降低,原子的自由运动减弱,物体能量降低。只要在凝结温度附近,使温度下降足够慢,原子排列就越来越规整,而形成结晶,这一过程称为退火过程。物体的上述结晶过程可对应多变量函数的优化过程,因此可以模拟退火过程而研究多变量的优化问题。

虽然 Hopfield 网络的状态收敛于所给的能量函数极小点的某一状态,但是,因为在 Boltzmann 机的状态迁移中引入了概率,故网络各状态收敛于各自出现的概率状态(平衡状态)。收敛于平衡状态后,再以相同的时间发展规则反复进行,其中计算各状态出现的次数的时间平均值,即可求出各状态出现的概率分布。此概率分布叫做 Boltzmann 分布。 Boltzmann 分布函数  $Q(E_n)$  的显著特征是:它仅是状态能量的函数。其表达式如下:

$$Q(E_n) = \left( \frac{1}{Z} \right) \exp \left( \frac{-E_n}{T} \right) \quad (38.5-68)$$

$$Z = \sum_n \exp \left( \frac{-E_n}{T} \right) \quad (38.5-69)$$

式中  $Z$ ——为了概率分布正规化的系数;

$T$ ——状态迁移规则(式 38.5-65)中的温度参数;

$E_n$ ——状态  $n$  的能量。

由此式可知,能量越低,实现概率越大,即“以最大概率出现最小能量状态”。

下面我们来考察温度参数  $T$ ,从式(38.5-68)可知,相对于实现概率  $Q(E_n)$ ,温度越高,其越不敏感;温度越低,其越敏感。当温度  $T \rightarrow 0$  时,最小能量状态的实现概率为 1,其他状态的概率全为 0。因此如果有效地利用此特性,就可能使网络的状态总是收敛于能量函数的最小点。可是,值得注意的是,正如我们在前面所讲的那样, Boltzmann 机的状态迁移规则,在温度  $T \rightarrow 0$  的极限时,与 Hopfield 网络是一致的。所以如果从

最初开始就让温度  $T = 0$ ，按时间发展规则，其状态会收敛于极小点，未必会收敛于最小点。

因此，最初以高温度出发，让网络按时间发展，在到达平衡点以后，保证平衡状态不破坏的条件下，慢慢地降低温度，最终使温度降低到 0 的极限的方法，叫做模拟退火法 (simulated Annealing)。这种方法，类似于把金属材料通过加热，再慢慢地冷却，以消除其内部的缺陷的退火原理。

模拟退火要点是温度下降的幅度。假如很快地降低温度，仍会收敛于极小点。关于这个问题，S. Geman 和 D. Geman 提出，如果温度  $T(t)$  下降满足式 (38.5-70) 的条件，网络一定收敛于最小点。

$$T(t) \geq c/\log(1+t) \quad (38.5-70)$$

式中  $t$ —— Boltzmann 机的时间发展规则的使用次数；

$c$ ——常数。

上式虽然由模拟退火保证了收敛于最小点，但由于式 (38.5-70) 是取对数的形式，温度变化非常缓慢，计算需要相当长的时间，因此，使用模拟退火时如何提高其计算速度乃是关键。

### (3) Boltzmann 机模型的学习算法

以 Boltzmann 机的时间发展规则动作，其状态的出现概率收敛于 Boltzmann 分布所表示的平衡状态。处于平衡状态各状态的出现概率，是根据 Boltzmann 分布，由状态的能量值而得到的。因为确定各状态能量值的能量函数是由神经元间的连接权值和神经元的阈值等网络参数所决定的，所以通过适当调整这些参数，可以实现所期望的状态的出现概率的平衡分布。而这里的调整就相当于 Boltzmann 机的学习。

实际上，Boltzmann 机的全部神经元可分为可视单元群 (visible units) 和隐单元群 (hidden units) 两个单元群 (如图 38.5-27a 所示)。进行学习是为了使可视单元群的状态的平衡分布与所期望的概率分布相一致。作为特例，如果隐单元的个数为 0，则进行所有神经元的学习。学习的方法分为自想起型学习和相互想起型学习两种。

自想起型学习是对学习时所提出的目标分布 (相对于网络来讲，是对应可视单元的外部环境) 实现模拟式学习的过程。学习结果是用网络参数表达外部环境的概率结构。

相互想起型学习是将可视单元群分为输入单元群和输出单元群 (如图 38.5-27b 所示)，在将输入单元群的状态 (输出值的群) 固定时，使输出单元群的平衡分布与期望的概率分布相一致。Boltzmann 机通过学习，建立起输入单元群 (输入模式) 和输出单元群 (输出模

式) 之间的有条件概率关系，这种相互想起型学习可以实现联想记忆的功能。例如，在把输入单元群固定表达为苹果时，假定在输出单元群中与之相对应，以概率的形式出现“红色”、“圆的”等模式，那么，网络就会从苹果联想出“红色”、“圆的”。

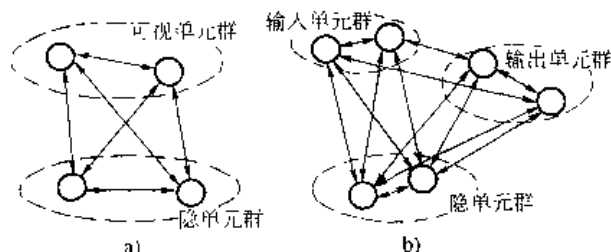


图 38.5-27 自想起型学习和相互

想起型学习的 Boltzmann 机

a) 自想起型 b) 相互想起型

下面介绍 Boltzmann 机的自想起型学习方法。将其扩张易于类推出相互想起型学习，限于篇幅，请读者参考文献“Boltzmann 机器的学习算法”。

假设可视单元群的状态为  $V_a$  (可视单元数的矢量)，隐单元群的状态为  $H_b$ ，其中下角标  $a, b$  分别为其状态的号码。想让 Boltzmann 机学习的可视单元群的期望分布函数为  $P(V_a)$ ；网络所有单元的平衡分布为  $Q_\omega(V_a, H_b)$ ，其中下角标  $\omega$  代表性地表示单元间的连接权值和单元的阈值等参数；与指定的状态  $(V_a, H_b)$  相对应的状态能量用  $E_\omega(V_a, H_b)$  表示的情况下，其  $Q_\omega(V_a, H_b)$  用下式的 Boltzmann 分布给出：

$$Q_\omega(V_a, H_b) = \left[ \frac{1}{Z} \exp\left\{ -\frac{E_\omega(V_a, H_b)}{T} \right\} \right] \quad (38.5-71)$$

$$Z = \sum_{a,b} \left[ \exp\left\{ -\frac{E_\omega(V_a, H_b)}{T} \right\} \right] \quad (38.5-72)$$

故仅考虑可视单元群的情况下，其状态的分布函数  $Q_\omega(V_a)$  为

$$Q_\omega(V_a) = \sum_b Q_\omega(V_a, H_b) \quad (38.5-73)$$

此时，期望分布函数为  $P(V_a)$  与网络实现的所有单元的平衡分布  $Q_\omega(V_a)$  间的差异为

$$G(\omega) = \sum_a P(V_a) \log \left\{ \frac{P(V_a)}{Q_\omega(V_a)} \right\} \quad (38.5-74)$$

其中： $G(\omega)$  称为 Kullback 信息量 (Kullback-Leibler divergence)，表示分布函数间的距离，无论对于怎样的分布  $Q_\omega(V_a)$ ，均有  $G(\omega) \geq 0$ ，等号仅在  $P(V_a)$  和  $Q_\omega(V_a)$  完全一致时成立。

把  $G(\omega)$  作为学习的评价函数，假如使  $G(\omega)$  趋向很小时确定网络的参数  $\omega$ ，就能以  $P(V_a)$  近似地构成实行网络。

仅让  $\omega$  的微小量  $\delta\omega$  变化的情况下，因为  $G(\omega)$  变



化为  $G(\omega + \delta\omega)$  可用下式给出:

$$G(\omega + \delta\omega) = G(\omega) - \delta\omega \cdot \left\{ \frac{\partial G}{\partial \omega} \right\} \quad (38.5-75)$$

所以,把  $\epsilon$  假设为一微小常数,把  $\delta\omega$  用下式表示:

$$\delta\omega = -\epsilon \left\{ \frac{\partial G}{\partial \omega} \right\} \quad (38.5-76)$$

则  $G(\omega)$  必定能单调减小,即有

$$G(\omega + \delta\omega) \leq G(\omega) \quad (38.5-77)$$

如此,开始就选择适当的初始值  $\omega$ ,用式(38.5-76)反复地修正  $\omega$ ,能够得到满足期望分布(使  $G(\omega)$  为极小值)的参数。式(38.5-76)中的微分,由下式给出:

$$\left( \frac{\partial G}{\partial \omega_{ij}} \right) = - \left( \frac{1}{T} \right) (p_{ij}^{(+)} - p_{ij}^{(-)}) \quad (38.5-78)$$

$$p_{ij}^{(+)} = \sum_a P(V_a) \frac{\sum_b S_i S_j \exp[-E_w(V_a, H_b)/T]}{\sum_b \exp[-E_w(V_a, H_b)/T]} \quad (38.5-79)$$

$$p_{ij}^{(-)} = (1/Z) \sum_{a,b} S_i S_j \exp[-E_w(V_a, H_b)/T] \quad (38.5-80)$$

$p_{ij}^{(+)}$  是先把可视单元群的输出值按照期望分布  $P(V_a)$  给以固定,仅让剩余的隐单元群的单元按照 Boltzmann 机的时间发展规则动作,在达到平衡状态时,单元  $i$  和  $j$  的输出值同时为 1 的期待值( $S_i$  和  $S_j$  间相关)。 $p_{ij}^{(-)}$  是让网络所有单元(可视单元群和隐单元群)自由,在达到平衡状态时,单元  $i$  和  $j$  的输出值同时为 1 的期待值。把以上的式子代入式(38.5-76),可以得到单元间连接权值的修正量表达式:

$$\delta\omega = -(\epsilon/T)(p_{ij}^{(+)} - p_{ij}^{(-)}) \quad (38.5-81)$$

上式中右边的第一项,是表示与单元  $i$  和  $j$  的输出值的相关成比例使其连接权值增大的项,可见这与 Hebb 的学习规则相类似。右边的第二项表示与单元  $i$  和  $j$  的输出值的相关成比例使其连接权值减小的项,也叫做反学习(unlearning)项。总之,网络在通过可视单元群同外部环境相连接时(醒时)进行学习;同外部环境连接切断时(睡眠时)进行反学习。可见 Boltzmann 机的学习理论有许多都是同睡眠和记忆(学习)的过程相对比进行的。

在式(38.5-78)中,虽只表示了相对单元间连接权值  $\omega_{ij}$  的微分,可是相对于各单元阈值  $\theta_i$  的微分,假如把阈值  $\theta_i$  看作为输出值总是为 1 的假想单元同单元  $i$  间的单元连接权值,那么其处理能以  $\omega_{ij}$  的相同形式进行。以下,将以上的学习规则作一整理。

1) 以概率  $P(V_a)$  将可视单元群的输出值固定为  $V_a$ 。

2) 把隐单元群的状态设定为温度为  $T$  的平衡状

态。

3) 统计所有的单元对(单元  $i$  同单元  $j$ ) 在平衡状态,同时输出为 1 的次数  $n_{ij}^{(+)}$ 。

4) 让可视单元群自由,所有单元的状态到达温度为  $T$  的平衡状态。

5) 统计所有的单元对(单元  $i$  同单元  $j$ ) 在平衡状态,同时输出为 0 的次数  $n_{ij}^{(-)}$ 。

6) 1) ~ 5) 反复进行,将算出的  $n_{ij}^{(+)}$  和  $n_{ij}^{(-)}$  的各自的平均值作为  $p_{ij}^{(+)}$  和  $p_{ij}^{(-)}$ 。

7) 所有单元间的连接权值  $\omega_{ij}$  按下式进行修正:

$$\delta\omega = -(\epsilon/T)(p_{ij}^{(+)} - p_{ij}^{(-)})$$

8) 返回 1)。

在以上学习规则的 2) 和 4) 中,为了实现温度  $T$  的平衡状态,虽然是根据 Boltzmann 机的时间发展规则进行计算的,但此时是由比  $T$  更高的温度开始计算,按照模拟退火法一边冷却,一边向温度  $T$  的平衡状态迁移,实现到稳定的温度  $T$  的平衡状态。

## 2 非线性振动的自学习建模

在计算机集成技术迅速发展的要求下,简单且易于操作的非线性系统识别技术的研究和开发既具有客观的迫切性,也具有广泛的应用前景。有关利用神经网络进行非线性的系统识别的研究虽已有许多,但是以简单的神经网络结构和简单的学习,能够高精度、大泛用地预测出其时序列非线性响应的研究仍存在着许多问题。例如,在生物或机械等系统建模中,只允许或只能通过简单加振测定对象的响应,并需依此进行其未知非线性系统识别。由于对象的实验条件局限性,非线性特性的未知性,以及可使用学习信息的有限性,常常难以确定神经网络的结构、输入和输出的构成以及选用的学习信息。因此,把非线性振动的简单响应作为学习信息,以实现系统时序列非线性响应预测是生物、工程等领域中通过试验完成系统建模所需求的重要手段。

本节从工程应用的实际出发,列举一适用于非线性振动系统识别的神经网络模式。其基本思想是,从神经网络的预测结构形式、输入和输出的构成以及有限学习信息的充分利用三个方面入手,以尽可能少且简单的学习信息,实现非线性系统的高精度、大泛用性时序列预测。文中,通过非线性振动时序列脉冲响应的学习,实现强制激振响应的预测,通过时序列周期响应的学习实现混沌响应的预测,验证了所介绍的神经网络模型的有效性。

### 2.1 神经网络和系统识别

(1) 识别的对象和系统预测的条件

研讨的非线性振动解析模型为

$$m\ddot{x} + cxF_R(x, t) = F(t) \quad (38.5-82)$$

其中:  $x(0) = x_0, \dot{x}(0) = \dot{x}_0$  为初始条件;  $F_R(x, t)$  为非线性复原力;  $\ddot{x}(t), \dot{x}(t), x(t)$  分别为系统的加速度、速度和位移。激振力  $F(t)$  为脉冲力时, 使用如下的半波正弦形式:

$$F(t) = \begin{cases} B\sin(\omega t) & (0 \leq t \leq t_m) \\ 0 & (t > t_m) \end{cases} \quad (38.5-83)$$

其中:  $t_m = \pi/\omega$ , 激振力  $F(t)$  为强制激振时, 使用如下的正弦形式:

$$F(t) = B\sin(\omega t) \quad (38.5-84)$$

以少量、简单条件下的时序列响应作为教师数据, 学习后的神经网络, 相对于任意的初始条件  $(x_0, \dot{x}_0)$  和任意的激振力(激振力的形式以及  $B, \omega, t_m$ ), 能够预测出其时序列响应。

(2) 提高神经网络预测精度和泛用性的方法

这里, 所追求的神经网络主要功能是在仅使用有限脉冲激振力和其响应的前提下, 神经网络通过这些有限信息的学习、记忆、联想, 达到对系统非线性特性的识别。在识别中, 神经网络学习的系统信息与需要预测的系统响应, 由于激振力、激振频率、初始条件的不同, 虽然在其响应上是不同的, 但是其确定性物理模型决定了其物理机理的同一性。这种非线性机理同一性用于神经网络表达, 不仅需要通过神经网络学习算法的改进来实现, 同时需要研究神经网络的结构、输入和输出的构成以及学习信息的有效利用来提高神经网络的预测精度和泛用性。

1) 确定神经网络的输入、输出的构成 在确定模型的离散时间域中, 假如给定  $t$  时刻系统的  $x(t), \dot{x}(t), F(t), F(t + \Delta t)$ , 那么  $\Delta t$  后系统响应的增加量  $\Delta x(t), \Delta \dot{x}(t)$  应该是确定的。因此, 把神经网络的输入确定为  $x(t), \dot{x}(t), F(t), F(t + \Delta t)$ , 输出确定为  $\Delta x(t), \Delta \dot{x}(t)$ , 通过学习其非线性机理和特性是有可能实现系统识别的。完成学习后的神经网络, 在学习信息条件附近的某一域内, 预测的初始值、强制激振力和频率的大小等取任意值作为已知条件, 从初始条件  $x(t), \dot{x}(t)$  开始预测。将其输出  $\Delta x(t), \Delta \dot{x}(t)$  代入式 (38.5-85)、式 (38.5-86)。

$$x(t + \Delta t) = x(t) + \Delta x(t) \quad (38.5-85)$$

$$\dot{x}(t + \Delta t) = \dot{x}(t) + \Delta \dot{x}(t) \quad (38.5-86)$$

把求出的值  $x(t + \Delta t), \dot{x}(t + \Delta t)$  再作为下一  $\Delta t$  后的系统输入, 再预测出  $\Delta x(t + 2\Delta t), \Delta \dot{x}(t + 2\Delta t)$ 。以此方式反复进行, 即可实现时序列  $x(t + \Delta t), \dot{x}(t + \Delta t)$  响应的系统预测。

2) 教师数据的结构及其矢量形式 在连续时间序列离散化为  $t = \Delta t \cdot k$  的情况下, 把教师数据条件  $(x(0), \dot{x}(0), B, \omega, t_m)$  下得到的系统时序列数据表示

为以下形式:

$$\tilde{x}_k = x(t) \quad (38.5-87)$$

$$\dot{\tilde{x}}_k = \dot{x}(t) \quad (38.5-88)$$

$$\tilde{F}_k = F(t) \quad (38.5-89)$$

$$\Delta \tilde{x}_k = \Delta x(t) \quad (38.5-90)$$

$$\Delta \dot{\tilde{x}}_k = \Delta \dot{x}(t) \quad (38.5-91)$$

图 38.5-28 所示的学习用神经网络的输入、输出的矢量关系由下式定义:

$$I_k = [\tilde{x}_k, \dot{\tilde{x}}_k, \tilde{F}_k, F_k]^T \quad (38.5-92)$$

$$\hat{O}_k = [\Delta \tilde{x}_k, \Delta \dot{\tilde{x}}_k]^T \quad (38.5-93)$$

通过教师数据的学习, 为了使神经网络具有非线性系统的频率、振幅和初始条件的依存特性, 必须提高神经网络的联想功能。为此, 有时需要用复数组试验条件

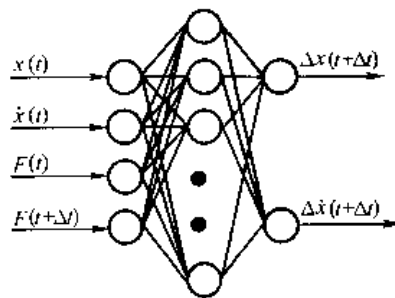


图 38.5-28 学习用三层 BP 网络

$(F^i, t_m; i = 1, 2, \dots, m)$  下的试验数据, 以增加学习信息, 增加对系统非线性机理的联想能力, 即把取得的复数组试验条件下的离散化时系列数据用下式的输入、输出矢量的序列形式表示:

$$\tilde{I} = \{\tilde{I}_1^1, \tilde{I}_2^1, \dots, \tilde{I}_n^1, \dots, \tilde{I}_1^m, \tilde{I}_2^m, \dots, \tilde{I}_n^m\}^T \quad (38.5-94)$$

$$\hat{O} = \{\hat{O}_1^1, \hat{O}_2^1, \dots, \hat{O}_n^1, \dots, \hat{O}_1^m, \hat{O}_2^m, \dots, \hat{O}_n^m\}^T \quad (38.5-95)$$

在预测中, 完成学习后的神经网络, 在给定预测条件的激振外力离散值和初始值下, 以图 38.5-29 所示的预测结构, 依次得到响应  $x_{k-1}, \dot{x}_{k+1}; (K = 1, 2, \dots, N)$ 。输入、输出的矢量关系通过 BP 学习法进行。

3) 神经网络的结构 对于时序列系统预测, 一般多使用反馈型神经网络结构。这里, 在特定的识别问题和识别条件下, 为了提高对非线性振动系统预测的精度和泛用性, 采用图 38.5-28 所示的多层型神经网络, 并且, 针对多层型神经网络的学习特征, 介绍以上的输入、输出的构成, 其神经元的输入、输出变换函数采用 sigmoid 函数模型。

下面通过两个非线性振动的解析例来验证所介绍的神经网络的预测结构和输入、输出的构成,以及有限学习信息的利用方法的有效性。为了证明方法的有效性,事例中物理模型以已知非线性振动模型为例。教师数据以及与预测结果相比较的基准数据,全由 RKG (Runge-Kutta-Gill) 法解析得到,以具有可比性。

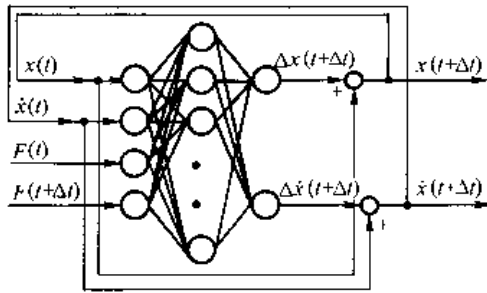


图 38.5-29 预测用网络的输入和输出的关系

### 2.2 非线性振动脉冲响应的学习和系统预测

通常,1 组试验条件下的时序列脉冲响应数据,其所持有的系统信息过于单调。事例中的学习数据,将 2 组试验条件下的时序列脉冲响应数据加工为 4 组,以尽可能减少学习信息。本节,叙述通过分段线性振动模型时序列脉冲响应的学习,实现其非线性系统响应预测的一仿真例。

#### (1) 分段线性振动的运动方程

式 (38.5-82) 中的非线性复原力由

$$F_R(x) = \begin{cases} (k + k_1)x + k_1\delta_1 & (x < -\delta_1) \\ kx & (-\delta_1 \leq x \leq \delta_2) \\ (k + k_2)x + k_2\delta_2 & (x > \delta_2) \end{cases} \quad (38.5-96)$$

所示的分段线性项表示。其运动方程表现了机械式制动鼓的物理模型。运动方程中的参数如表 38.5-3 所示。

#### (2) 分段线性振动脉冲响应的学习

学习方法使用 BP 法的修正力矩法。学习中,神经网络的输入层、中间层、输出层的单元数为 4、16、2。时间离散化增量为  $\Delta t = 2\pi\omega_n/L_B = 0.001s$ , 其中  $\omega_n$  为线性固有圆频率;  $L_B$  为线性固有周期的离散数;  $\eta = \omega/\omega_n$ ;  $\omega_n = \sqrt{k/m}$ 。

表 38.5-3 分段线性振动的非线性复原力参数

<i>i</i>	$k_i$ / (N/m)	$\delta_i$ / mm	$m$ / kg	
1	4000	1	$c$ / (Ns/m)	40
2	4000	-1	$k$ / (N/m)	0

教师数据使用 4 组 ( $m=4$ ) 条件为表 38.5-4 所示的半波正弦激振力下的时序列数据作为输入、输出矢量序列 (见图 38.5-30)。学习次数  $L_n=50000$ 。

表 38.5-4 分段线性振动脉冲响应学习的教师数据条件

<i>i</i>	1	2	3	4	$L_B$	200
$B_i/N$	20	20	-20	-20	$x_i, i=1, \dots, 4$	0
$t_{\omega_i}/s$	0.03	0.05	0.03	0.05	$x_i^*, i=1, \dots, 4$	0

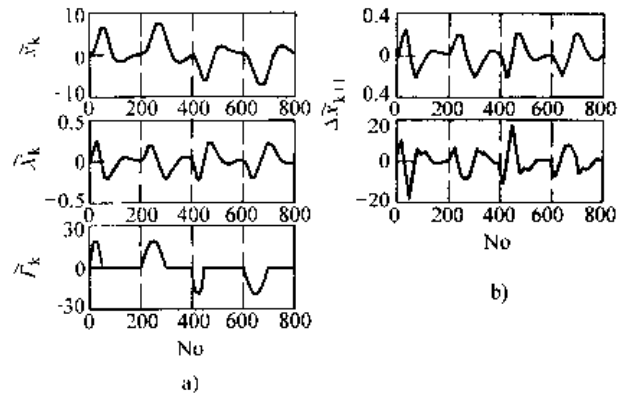


图 38.5-30 分段线性振动的 4 组 ( $m=4$ ) 学习数据

a) 输入数据 b) 教师数据

#### (3) 分段线性振动系统响应的预测

相对于学习的脉冲响应 2 组条件 ( $P, t_m$ ) 进行预测,其预测例如图 38.5-31a、b 所示;相对于强制振动响应 2 组条件 ( $\eta, P, x_0, \dot{x}_0$ ) 进行预测,其预测例如图 38.5-32a、b 所示。

### 2.3 Duffing 振动的学习和预测

#### (1) Duffing 振动的运动方程及解析

运动方程的物理模型为一端固定梁的大振幅振动。式 (38.5-82) 运动方程的非线性复原力项为

$$F_R(x) = k(x + \beta x^3) \quad (38.5-97)$$

运动方程中的有关参数如表 38.5-5 所示。

表 38.5-5 Duffing 方程的有关参数

$m$ / kg	1.0	$k$ / (Ns/m)	0.35
$c$ / (N/m)	0.0	$\beta$ / (N/m <sup>3</sup> )	1.0

#### (2) 学习和学习数据

根据前述的学习输入、输出关系,以图 38.5-33 中所示  $\square$  位置 ( $\omega=1, B=10.0, x_0=0.8, \dot{x}_0=0.0$ ) 为学习条件,即如图 38.5-34 所示,用 4 个周期 ( $m=1$ , 由 0 开始 200 个,  $\Delta t=2\pi/(50\omega)$ ) 的强制激振力和其时序列周期响应为输入,用其时序列响应的增量为教师数据。

从图 38.5-33 知,作为系统识别对象的 Duffing 模型,其解的特性与激振力和初值呈极强的敏感性。神经网络的学习参数由学习的平均误差和最大误差同时稳定收敛为基准决定,学习次数为 1000 次,学习误差如图 38.5-35 所示。与学习条件相同的响应预测结果如图 38.5-36 所示。由 RKG 法解析所得 3 个周期解的

响应相比较, 时序列响应、相平面轨迹 (Trajectory) 以及庞加莱映射 (Poincare map) 均显示出相当好的一致

性, 说明完成了高精度的学习。

(3) 由周期响应的学习实现系统响应的预测

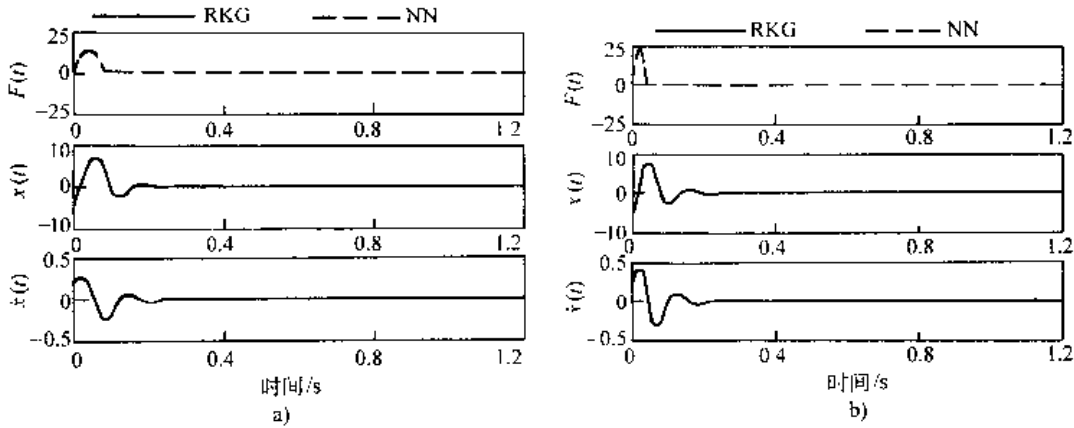


图 38.5-31 分段线性振动脉冲响应的预测例

a)  $B=15.0, t_m=0.08$  b)  $B=25.0, t_m=0.04$

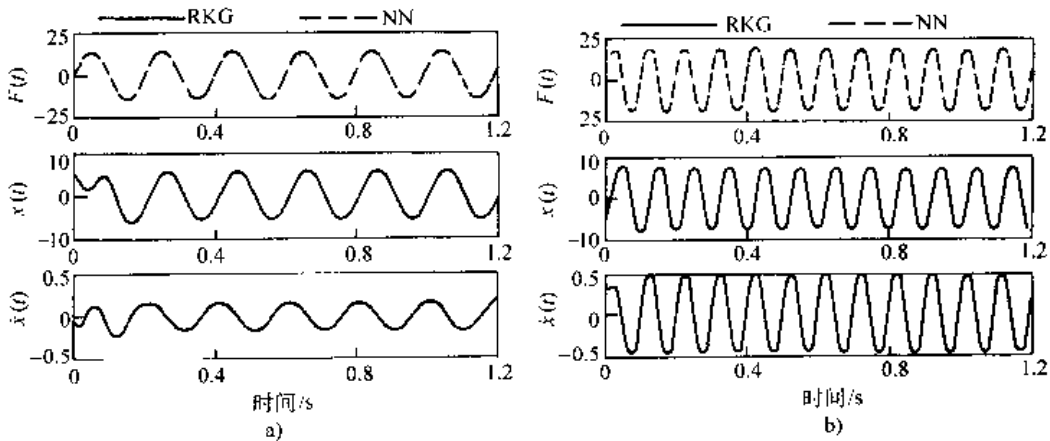


图 38.5-32 分段线性振动强制振动响应的预测例

a)  $\eta=0.5, B=15.0, x_0=0.005, \dot{x}_0=0.0$  b)  $\eta=1.0, B=20.0, x_0=-0.005, \dot{x}_0=0.2$

运动方程中的有关参数如表 38.5-3 所示。由 Ueda 的结果, 式 38.5-82 的 Duffing 方程式的响应在  $c-B$  平面被分为单纯混沌响应域、混沌和周期响应的共存域以及周期响应域。例如, 假定  $x_0=0.0, \omega=1$ , 由 RKG 法在共存域中 ( $P=10.0, x_0=0.8$  附近) 的时序列响应如图 38.5-33 所示。图 38.5-33 在  $B-x_0$  平面把响应分为混沌响应 (网格), 3 个周期响应 (口) 表示 (以后的图中与此相同, 网格和口分别表示由 RKG 法解析得到的混沌响应, 3 个周期响应)。由图 38.5-33 可以看出, 此非线性方程对初期值和激振力的大小非常敏感, 其非线性机理十分复杂。

(4) 预测

在这里, 由条件  $B, x_0, \dot{x}_0, \omega$  的组合, 以神经网络的预测结果同 RKG 法得数值解析结果相比较, 检验神经网络的预测泛用性。但是, 式 (38.5-82) 的衰减系数  $c$  和预测离散时间增量  $\Delta t$  与学习条件相同。

在式 (38.5-82) 的 Duffing 方程的  $\omega=1$  的前提下,

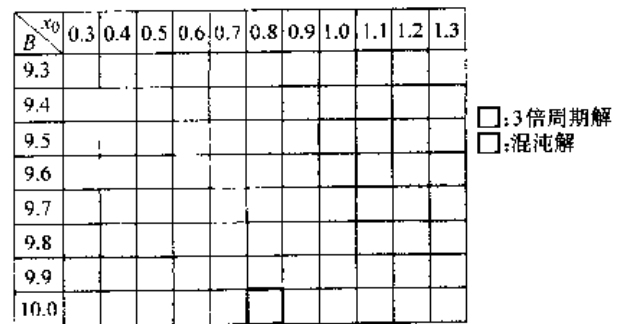


图 38.5-33 用  $B-x_0$  平面表示 Duffing 方程式的响应分类在学习条件的附近, 进行  $B-x_0, B-\dot{x}_0$  平面的预测。为了从多方面评价预测结果, 使用时序列响应、相平面轨迹以及庞加莱映射与 RKG 法的结果进行比较, 其响应的评价仅表示省去 4 周期的过渡响应后 40 周期的响应。

1)  $B-x_0$  平面的预测 图 38.5-36 定性地表示了  $\dot{x}_0=0.0$  条件下, 对应于  $B, x_0$  各条件的时序列响应预测结果同 RKG 法解析结果的比较。图 38.5-37 示出了

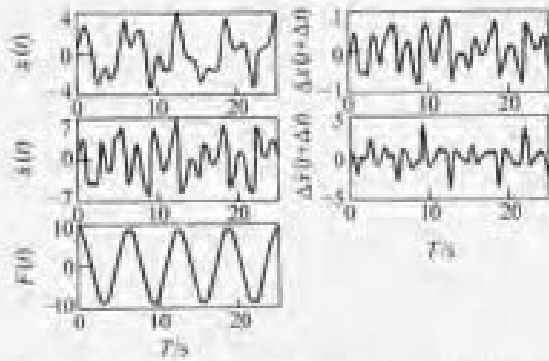


图 38.5.34 Duffing 非线性驱动 ( $\omega=1$ ) 学习数据

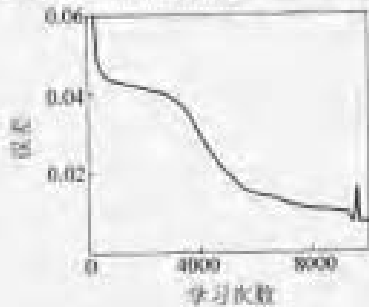


图 38.5.35 学习误差

从图 38.5.36 中 4 种评价表示 (⊙、○、○D、△) 的

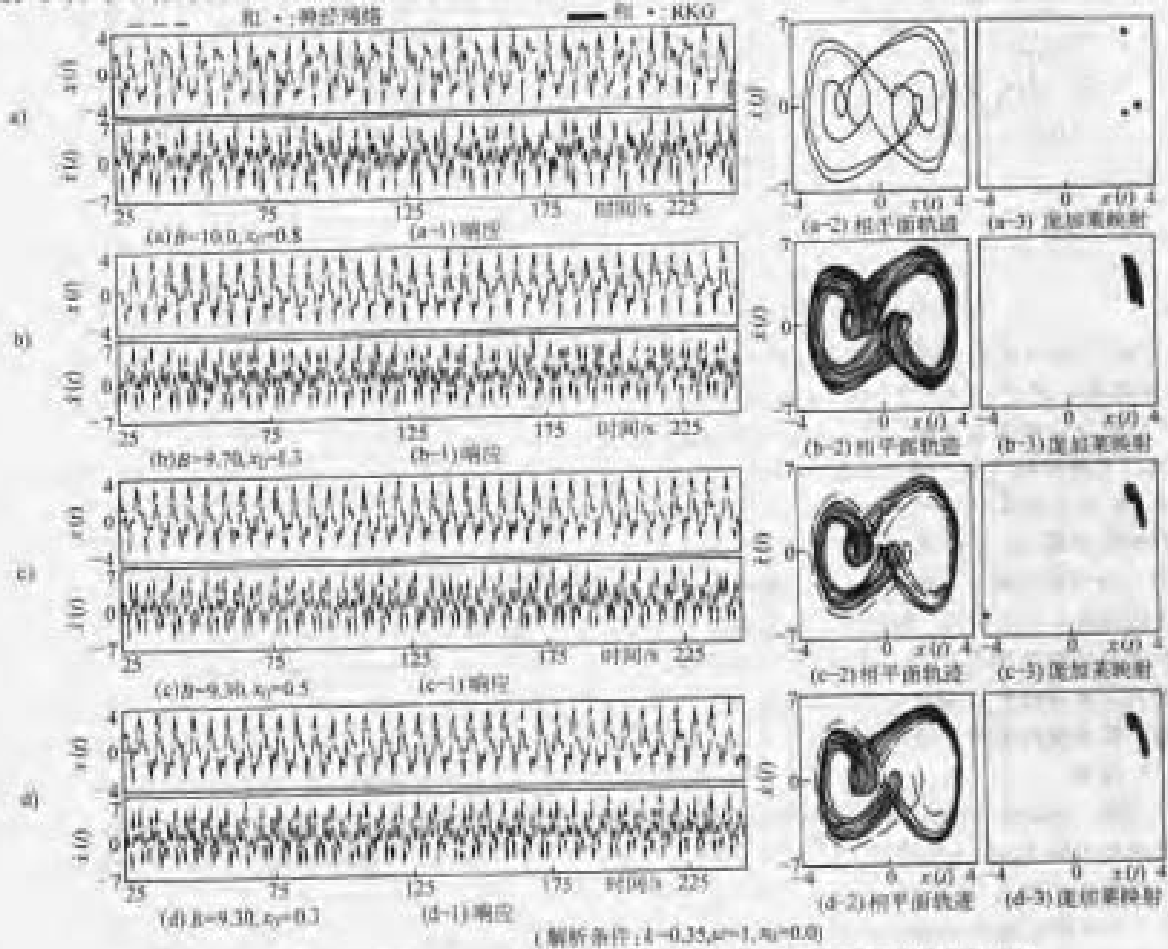


图 38.5.37 用  $B-x_2$  平面的预测条件预测的时序响应、相平面轨迹和庞加莱映射例

预测结果, 代表性地以 4 个预测例表示其时序响应、相平面轨迹以及庞加莱映射。图 38.5-37a 的预测条件与学习数据条件相同 (图 38.5-36 中附有  $\bullet$  记号之  $B, x_1$ ), 其预测结果与 RKG 法的结果一致, 用  $\odot$  表示。图 38.5-37b 的例中, RKG 法的解析结果是混沌响应, 神经网络的预测结果从时序波形和庞加莱映射看, 较好地预测了其混沌响应, 但在相平面轨迹出现偏差, 用  $\ominus$  表示。图 38.5-37c 的例中, 同样虽然定性地预测了混沌响应, 但在相平面轨迹上出现  $180^\circ$  误差, 用  $\circ D$  表示。图 38.5-37d 的例中, RKG 法的解析结果是 3 个周期的周期解, 但预测结果却是混沌响应, 用  $\triangle$  表示。

$B \backslash x_1$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3		
9.3	$\triangle$	$\circ$	$\odot$	$\odot$	$\odot$	$\odot D$	$\odot D$	$\odot$	$\odot$	$\odot$	$\odot$	$\triangle$	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; width: 10px; height: 10px; display: inline-block;"></div> 3 周期解</div> <div style="border: 1px solid black; width: 10px; height: 10px; display: inline-block;"></div> 混沌解

图 38.5.36 用  $B-x_1$  平面表示神经网络预测结果

图 38.5-37 中的记号  $\odot$ 、 $\circ$ 、 $\odot$ D、 $\triangle$ ，表示了神经网络的预测结果的精度。 $\odot$  表示时序列波形和相平面轨迹均呈现高的预测精度。 $\circ$  表示从相平面轨迹可看出其预测稍有误差。 $\odot$ D 表示实现了混沌响应的定性预测，但相平面轨迹上出现  $180^\circ$  误差。 $\triangle$  表示预测的定性结果与 RKG 法的解析结果截然相反（以下各图中  $\odot$ 、 $\circ$ 、 $\odot$ D、 $\triangle$  记号的意思均相同）。

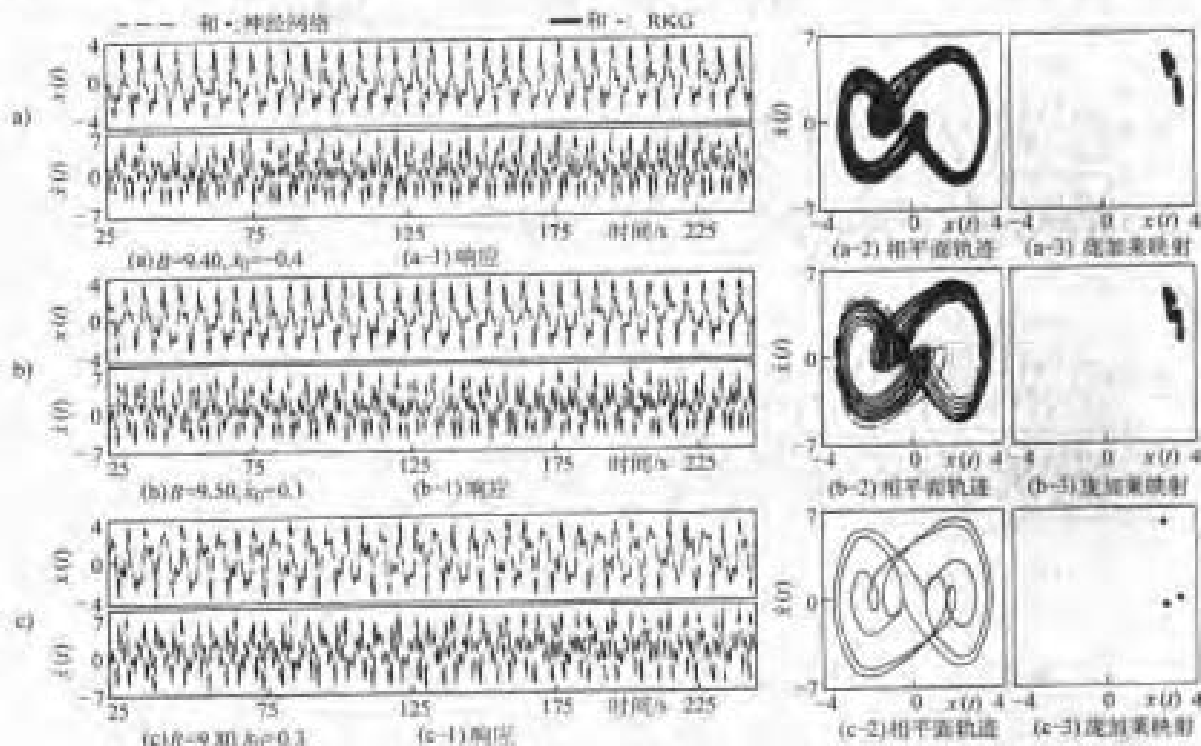
2)  $B-x_0$  平面的预测 图 38.5-38 定性地表示了在了  $x_0=0.8$  条件下，对应于  $B, x_1$  各条件的时序列响应预测结果同 RKG 法解析结果的比较。图 38.5-39 是图 38.5-38  $B-x_1$  平面中的预测例。

$B \setminus x_1$	0.5	0.4	0.3	0.2	0.1	0.0	0.1	0.2	0.3	0.4	0.5
9.2	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$ D	$\odot$
9.4	$\odot$ D	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
9.5	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
9.6	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$
9.7	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
9.8	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
9.9	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
10.0	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$

$\odot$  3倍周期解  
 $\odot$  混沌解

(解析条件,  $k=0.35, \alpha=1, x_0=0.8$ )

图 38.5-38 用  $B-x_1$  平面表示神经网络预测结果



(解析条件,  $k=0.35, \alpha=1, x_0=0.8$ )

图 38.5-39 用  $B-x_1$  平面的预测条件预测的时序列响应、相平面轨迹和庞加莱映射例

$B \setminus x_1$	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
0.4	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\triangle$	$\triangle$	$\triangle$	$\odot$	$\odot$
0.3	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.2	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.1	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.0	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.1	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.1	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
0.3	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$

$\odot$  3倍周期解  
 $\odot$  混沌解

(解析条件,  $k=0.35, \alpha=1, B=10.0$ )

图 38.5-40 用  $x_1-x_0$  平面表示神经网络预测结果

3)  $x_1-x_0$  平面的预测 图 38.5-40 定性地表示了

在  $B=10$  条件下，对应于  $x_1, x_0$  各条件的时序列响应预测结果同 RKG 法解析结果的比较。图 38.5-41 是图 38.5-40 中  $x_1-x_0$  平面中的预测例。

4) 相对激励频率变化的预测例 图 38.5-42 定性地表示了在了  $B=10.0, x_1=0.8, x_0=0.0$  条件下，对应于激励频率条件的变化，其时序列响应预测结果同 RKG 法解析结果的比较，但是，因图形表现的原因，其时序列响应的预测结果未必是省去 4 周期过渡响应的 40 周期。

### 2.4 预测精度和泛用性的考察

以上两例中，从预测结果同 RKG 法解析结果的一致性，充分说明了其预测精度；从所使用教师数据

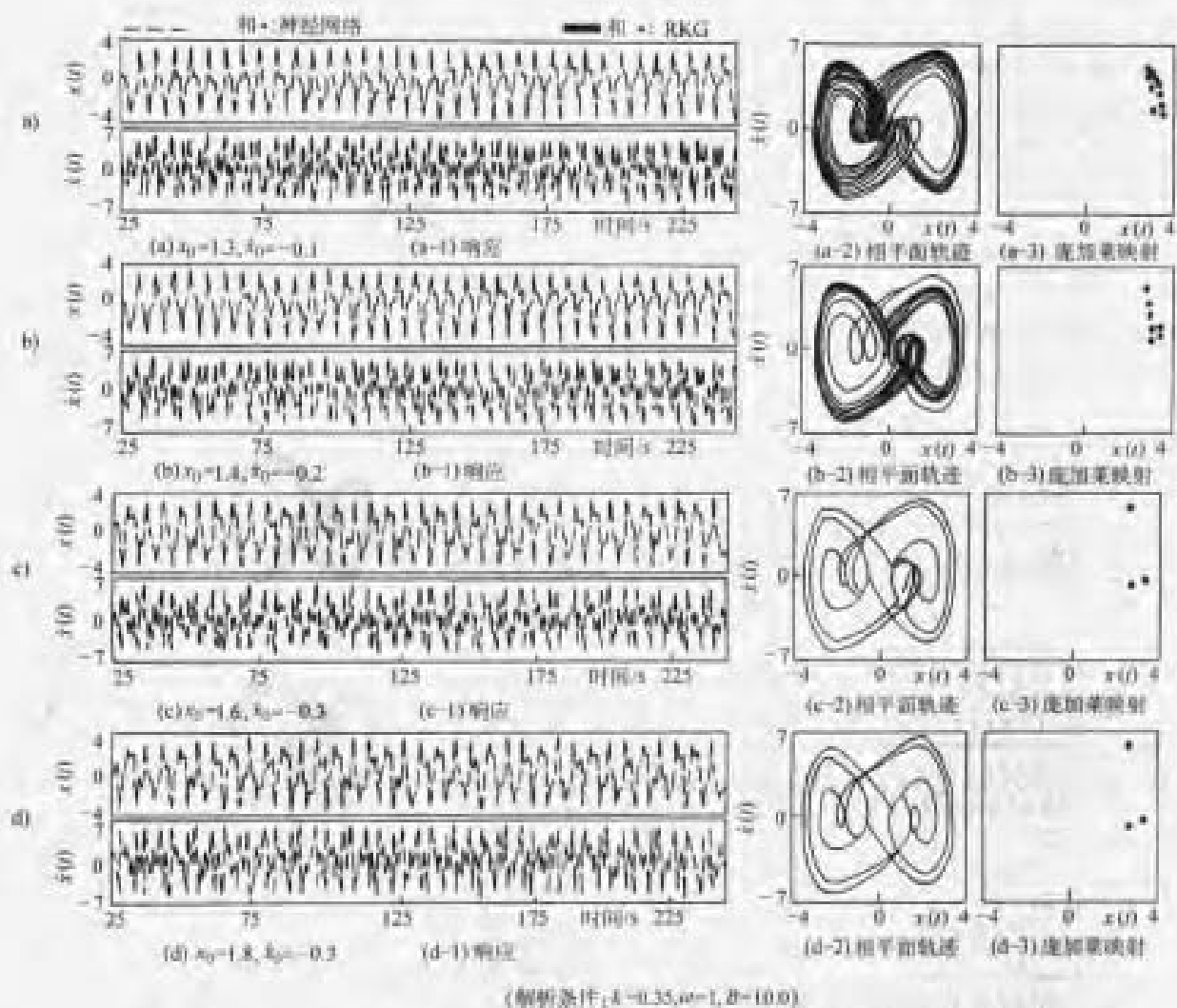


图 16-5-41 用  $x_0-x_1$  平面的预测条件预测的时序列响应、相平面轨迹和庞加莱映射

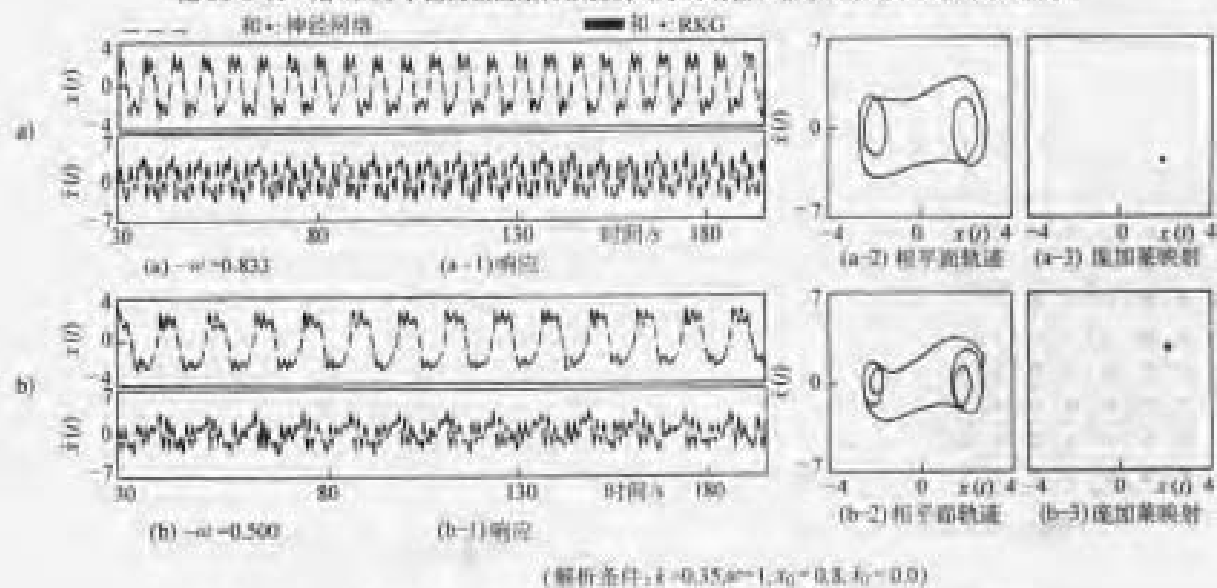


图 16-5-42 由预测参数  $\alpha$  变化得到的时序列响应、相平面轨迹和庞加莱映射

(脉冲或强制振动响应)的信息形式和信息量及系统预测时所给出的预测条件的多样性,充分说明了其预测具有相当的泛用性。

这里,说明了使用阶层型神经网络由简单的脉冲响应的学习,实现其复杂响应预测的可能性。通过对分段线性振动脉冲响应的学习,实现了对其时序非线性振动脉冲和强制响应的预测,通过在 Duffing 非线性振动的混沌响应和周期响应的共存域中,让神经网络学习其周期响应后,改变激励力的大小和频率以及初始条件,进行其响应的预测。预测的结果,用时序响应、相平面轨迹、庞加莱映射,与 RKG 数值积分的结果进行比较和判定。由此,验证了介绍的神经网络模型对非线性振动系统识别的有效性,即验证了所介绍的神经网络模型中的预测结构形式、输入和输出的构成以及有限学习信息利用方法的有效性。需要说明的是, Duffing 非线性振动中,在混沌响应和周期响应的境界等部分,由于其相对预测条件的敏感性,存在预测的难度。

### 3 基于学习的机械系统特性预测

#### 3.1 机械系统特性预测的问题

在机械结构设计的实际工作过程中,首先是进行方案设计,然后根据确定的结构方案,进行详细的技术设计,整个过程如图 38.5-43 所示。

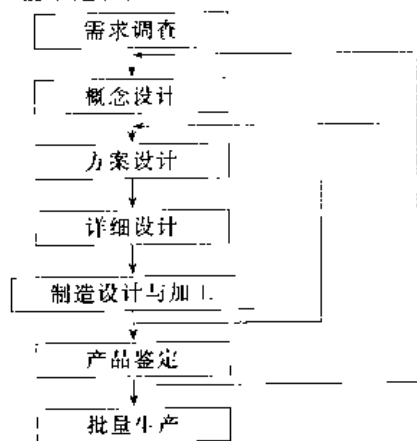


图 38.5-43 机械产品设计流程

对于机械结构方案设计者来说,在完成方案设计后,很期望对系统总体特性指标有一个比较可靠的估算,以便快速、有效地对结构方案设计参数进行更趋合理的改进和调整。在大系统、多品种机械产品设计中,建立起结构方案设计参数与结构总体系统特性指标之间的有机联系,是总体设计中的一大课题。一旦方案设计参数确定后,进入详细技术设计阶段,设计者将根据有关的技术规范、力学解析以及经验等进

行设计。假定在此阶段中,由于设计技术人员的不同,仅在结构局部会出现差异,并不会对系统特性带来大的影响的话,可以认为,方案设计所确定的各参数与系统特性指标之间存在着一种必然的对应关系。实际设计中,这种对应关系很难用一数学函数的方式予以表达。

一般来说,一个机械产品的方案设计着重于产品的功能需求,而产品的详细设计则着重于其功能实现的保障。机械结构系统的某些特性,在结构设计中,是需要极其重视并保证的参数。例如,根据机械结构工作精度或伺服控制等目标的要求,需要对机械结构系统总体的特性指标在结构方案设计阶段进行严格控制。

产品的详细设计一般遵从由机械结构设计到机械系统总体特性的校核计算,再回过头来对结构进行修正的循环的过程。一些小型简单的机械产品往往用力学建模,借助有限元方法或模态综合法就能较为精确地计算出结构系统的总体特性,如频率、声场或温度场等,甚至能够制作模型进行实验,以预测其总体特性。然而对于一些大型、结构参数非线性很强的结构系统,用以上的方法往往难以奏效。目前,一般只能根据经验来设计。也有根据以前的产品系列的大量数据用统计的方法获得未来产品结构总体特性的预测。而对于许多小批量小样本的产品,传统的统计方法则显得力不从心。为此,我们引入神经网络方法,用其本身的自学习功能和智能性,能够较好地解决强非线性和小样本问题。本节介绍在机械结构方案设计之后,详细技术设计之前,以已知机械结构方案设计参数作为基本条件,用神经网络方法预测出任意机械结构系统特性参数的思想和算例。这对在机械系统设计中,建立智能化构筑系统特性预测方法的研究是有益的。

#### 3.2 机械系统特性预测的基本模型

机械结构系统受零部件结构、体积、重量以及实现其功能的连接方式和形式等诸多因素的影响。从传统动力学观点出发来建模,实现系统特性的预测,往往存在以下困难:

1) 在初步确定了方案设计参数,但并无详细的结构技术资料的前提下,难以得到其较为精确的力学模型。

2) 机械结构方案设计参数多,且其参数对结构系统特性指标影响复杂,无法建立适合计算任意机械结构系统特性的数学模型。

3) 即使已知机械结构系统各零部件单元对整体结构特性的影响,且已知各个单元详细的技术设计,但由于零部件连接有着强非线性因素影响,严格地求解和



推导出其系统特性也是相当困难的。

假如我们承认方案设计参数与系统特性之间,由现有详细设计技术建立起了一种必然的对应关系,那么,现有的各类机械产品,则内含着其详细设计过程的技术规范、力学解析以及经验等知识和信息。随着技术、材料的进步,将来的产品也将内含其详细设计过程中的新知识和新信息。利用这些现有机械结构的内含知识和信息,建立起其方案设计参数与系统特性之间的联系,是机械系统特性预测基本模型的立足点。

虽然在方案设计阶段,用数学的方法无法建立方案设计参数和系统特性之间的表达,但可以将它想象成一个黑盒,其系统特性预测的基本关系如图 38.5-44 所示。利用现有产品的内含知识和信息构筑这样一个模型来模拟黑盒,使它基本符合某一产品系列个例。如果将产品系列的某一个产品方案设计参数输入模型,能够得到忠实其产品实测特性的系统预测特性,则可为方案设计提供定量的反馈评价信息。我们采用具有学习功能的人工神经网络(ANN)方法来构筑机械结构特性这个黑盒的模型。



图 38.5-44 系统特性预测的基本关系

### 3.3 雷达结构系统固频的预测例

雷达结构系统的固有频率(一阶振频)在雷达结构设计中,是需要极其重视并保证的参数之一。根据雷达跟踪精度及伺服控制的要求,需要对这一雷达结构系统总体的特性指标在结构方案设计阶段进行严格控制。雷达结构系统中包含诸如静压油垫、轴承等多项非线性环节,这导致即使在系统模态综合中也很难精确地算出它的频率特性,而且雷达系列产品又属于典型的小样本产品,难以用统计学的方法建立其系统特性预测模型。

用人工神经网络预测雷达结构固有频率,即网络的输出信号为雷达结构的固有频率,网络的输入参数为雷达结构总体设计方案中给出的确定性参数。选择网络输入参数贯彻以下三个原则:

- 1) 这些参数反映雷达的天线-天线座系统的结构特性,并且是雷达结构总体方案论证书中指定的总体方案参数,而不是技术设计中的所有结构的具体参数。
- 2) 这些参数对雷达结构系统的固频产生最直接最显著的影响作用。
- 3) 这些参数互不相关,这样可以避免参数的隐性重复选择。

根据这三项原则,将网络的输入参数确定为以下八个:① $x_1$ ——天线口径(m);② $x_2$ ——天线重量(kg);③ $x_3$ ——方位转台支承类型(1.静压轴承 2.交叉滚子轴承 3.标准轴承);④ $x_4$ ——方位转台支承直径(m);⑤ $x_5$ ——俯仰支承类型(1.向心推力球轴承 2.交叉滚子轴承 3.圆锥圆柱滚子轴承);⑥ $x_6$ ——俯仰支承轴承数;⑦ $x_7$ ——俯仰支承轴承外径(mm);⑧ $x_8$ ——方位支臂位置(1.内置 2.外置)。输出学习参数为 1 个: $y$ ——结构系统的固有频率。

人工神经网络是由 8 个输入节点和 1 个输出节点构成的三层层状网络。中间隐含层节点定为 15 个。

#### (1) 神经网络的学习

我们获取的学习数据如表 38.5-6 所示,从表中 No. 1~8 的八组学习数据可以看出,雷达总体方案参数  $X$  对固频  $Y$  的影响呈现出非线性关系。在对表中的数据进行规范化处理后,经过 9000 次的学习过程,结果趋向收敛。其误差小于 0.01,相对误差小于 5%。

#### (2) 系统固频的预测

1) 未学习的雷达结构数据的预测 将待考察的雷达结构数据如表 38.5-7 所示。经网络回忆计算后得到固有频率值为 6.18,与实际的数据误差为从网络泛用性评价角度讲是有效的。

表 38.5-6 用 ANN 预测雷达结构固频的学习数据

No.	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$y$
1	2.5	90	3	0.29	1	2	170	1	18.0
2	3.6	150	2	1.29	2	2	940	1	15.0
3	4.2	500	1	2.15	2	2	340	2	9.0
4	5.0	1000	2	2.00	2	2	340	2	8.0
5	6.0	3800	2	2.20	2	2	350	2	7.3
6	7.0	3000	1	2.20	3	3	380	2	6.7
7	8.5	3500	1	2.50	3	2	400	2	6.8
8	10.0	4000	1	2.50	3	2	400	2	5.9

表 38.5-7 待考察的雷达结构预测输入数据和预测结果

说明	预测输入数据								预测固频	实测固频
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$		
数据	9.0	3500	1	2.45	3	3	380	2	6.18	6.00

2) 已学习的雷达结构数据的预测检验 在图 38.5-45 中,我们以表 38.5-6 中 No. 为序,分别以已学习过的数据  $x_1 \sim x_8$  输入网络,得到的预测结果用●表示,原网络输出学习信号用▲表示,以此来直观地检验学习的精度。从图中可以看出,系统的预测精度基本上忠实地反映了原学习数据。

综上,在机械产品系统设计的方案设计阶段,以方案设计参数为输入信号,用神经网络方法来预测机械

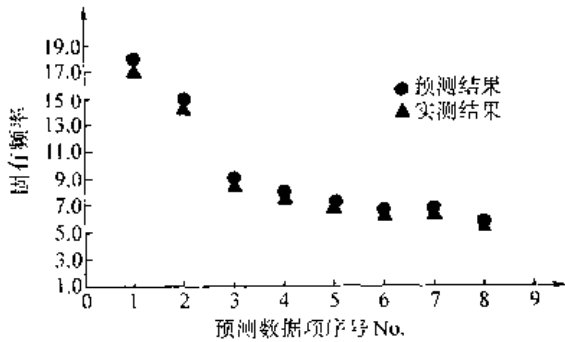


图 38.5-45 已学习的雷达结构数据的预测检验

系统的特性是机械系统智能化设计可借鉴的方法之一。其应用将具有以下特点：

① 为方案设计提供定量的反馈评价信息。

② 将现有产品的内含知识以信息集成的方式来辅助系统设计。

③ 方法简单,易于操作,可通过追加学习的方式来适应新信息的集成,因此具有实用价值。

本节介绍的思想和方法,虽然仅以雷达结构系统固有频率预测的实例作了仿真解析,但在机械系统设计中的应用具有普遍性。

#### 4 神经网络专家系统的智能设计体系结构

传统的 CAD 技术经过近 40 年的发展,已经在机械、建筑、电子、广告等诸多领域取得了巨大的成就。正是由于 CAD 技术的出现,才使工程设计人员从繁重的、枯燥的绘图劳动以及数值计算中解放出来的愿望变成了现实,但是,随着科学技术的不断进步,特别是计算机技术的飞速发展,人们对 CAD 技术也提出了更高的要求,希望 CAD 技术除了拥有原来的优势之外,还能够代替设计师的一部分智能活动,让计算机向设计师一样具有思维,能够自动进行设计,从而减少 CAD 过程中对人的依赖。这样传统的 CAD 系统有必要扩展为智能 CAD 系统。

智能 CAD 是 CAD 技术和人工智能技术的结合,也就是在原有的 CAD 系统中集成上知识处理系统。其目的是让计算机参与设计过程,从而将设计自动化引向深入,这是 CAD 技术在学术上深化的一个十分重要的方向,现在常见的,也是比较成熟的是 CAD 专家系统,许多领域的专家系统在实际应用中是十分成功的。

专家系统之所以能取得成功是因为专家系统的求解是建立在领域专家知识的基础上的,但同时也存在一些问题,比如任何领域的专家并不总是用规则来思考问题,在这种情况下,专家系统本身并没有真正模仿人的推理过程,而神经网络理论是借鉴真实神经系统

的某些功能,抽象、概括、简化而成的方法,是在现代神经科学研究成果的基础上提出的,它反映了人脑功能的若干特性。将神经网络技术应用于设计过程,对于处理设计条件描述不够充分的设计问题,具有极其重要的价值,应用前景十分广阔。神经网络专家系统的研究正是随着神经网络的发展而逐步发展起来的。

#### 4.1 建立神经网络专家系统的必要性

专家系统本身存在许多如知识表示、知识获取、处理大型复杂问题比较困难等理论和技术上问题,与之相同的神经网络虽然具有自学习、容错性、自适应性以及并行结构和并行处理能力等许多优点和特征,但也存在一些本身固有的缺点,比如网络连接模型表达的复杂性、训练过程的不稳定性、训练时间过长以及网络硬件实现技术难等等。而神经网络专家系统这种混合型专家系统正是综合利用神经网络方法和传统人工智能方法的长处而生成的专家系统,专家系统用来处理基于规则和事实的知识,进行逻辑推理;神经网络用来处理不充分、容易变化的知识,进行联想、分类和识别等形象思维。逻辑推理和形象思维是非常巧妙的相互配合而形成有机整体的,使基于人工智能的符号系统和基于神经网络技术的神经网络系统结合起来,充分发挥各自的优点,避免各自的不足,是一个应用前景十分广阔的专家系统模型结构。

#### 4.2 面向设计的智能平台

##### 4.2.1 专家系统和神经网络的结合方式

对于专家系统和神经网络的结合方式我们拟采用嵌入式结构,即神经网络作为一个模块嵌入专家系统中。其结构如图 38.5-46 所示。

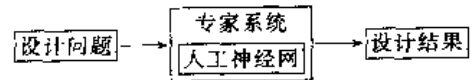


图 38.5-46 专家系统和神经网络的结合方式

图 38.5-46 中神经网络模块的作用是将基于神经网络模块收集的信息,转化为专家系统推理过程中所需的事实和规则。具体转化过程如图 38.5-47 所示。

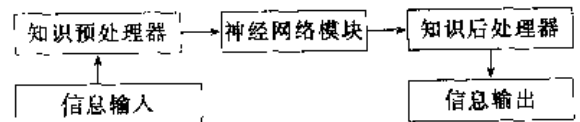


图 38.5-47 神经网络模块中知识的转化

##### 4.2.2 智能平台的“外壳”结构

我们所说的系统“外壳”结构是指专家系统中除知

识库以外的公共部分,整个体系包含知识库、推理机、全局数据库、人机交换接口、知识获取、解释机构、图形处理等部分。由于结合了神经网络技术,所以系统中还有神经网络模块。它们通常具有通用性,当我们进行具体设计时,只要将要设计问题知识按规定的知识库描述语言格式编辑成知识库,并将获取的知识进行检验,生成所需的专家系统,从而可以进行特定环境的设计活动。它一般由专家系统外壳和知识库生成与管理子系统两大部分组成。

智能平台体系结构的框架如图 38.5-48 所示。

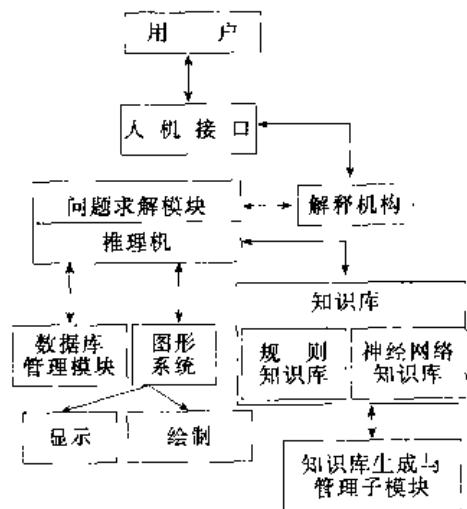


图 38.5 48 体系结构图

几点说明:

1) 理想的智能支撑平台应是基于知识的具有较高交互性和可视化程度的专家系统设计工具,它能提供各种知识表达方式、推理方式及图形支撑数据库、知识库管理系统以及灵活、友好的界面,以便对不同的设计问题进行有关的功能扩充,从而达到快速设计专家系统的目的。

2) 支撑平台中含有神经网络模块,以便提高利用神经网络学习任意复杂非线性映射关系的能力,掌握学习问题领域中难以明确表示的隐式知识,是一种比人工方法获取经验性知识更为自然和有效的方法。神经网络知识的获取不需要由知识工程师来整理、总结、消化领域专家的知识,只需用领域专家解决问题的事例或范例来训练网络,使在同样输入的条件下能够获得与专家给出的解答近可能的输出。

3) 数据库管理模块是对设计结果进行管理,并可以对设计信息查询、浏览等操作,并提供数据信息的交流、传输。

4) 推理机可以根据当前的状况,利用知识对问题领域的问题进行求解。

5) 知识库生成与管理子系统是本工具的重要部

分,它实现如下功能:

① 按特定的知识描述语言,通过人机交互,将整理好的知识存入知识库。

② 对建造完成的知识库进行结构与语法检查,显示出错信息,提示建造者进行修改。

③ 连接推理机对知识库进行动态模拟调试,进行完备性和一致性检验。

#### 4.2.3 设计求解过程

运用该系统进行设计的过程是先从设计专家处获取设计知识——包括原理性知识和专家和经验——装入知识库中,专家系统根据这些知识和设计要求、初始数据,启发式的探索求解通路,模拟专家的设计过程,在设计的各层次上,自动的判断做出决策,确定参数,最后直接给出一个或几个合理的目标设计方案供设计者选择,并以工程图的形式得到一个基本设计。

我们为此建立了一个设计求解的总模型:所需的数据来自全局数据库,所需的知识来自规则库和神经网络模块,当用户输入一定的信息后,按一定的策略对知识库进行推理,同时显示系统推理的结论,当结论正确时,用户可以继续以下的推理工作;当结论不符时,用户可以交互地修改某些参数;用户可以通过知识库中的知识确定结论正确与否。重复以上的过程,直到把整个求解完成为止,交图形系统显示或绘制。

#### 4.2.4 知识的处理方法

知识库是该体系的核心,我们可以将设计问题时遇到的知识库分成两类:一是关于设计过程的知识,即关于如何进行设计的知识,其中包括设计的一般原理和人类设计师的经验;二是关于设计对象的知识,即设计对象的部件、结构、材料、用途等。该体系的知识库由两部分组成:规则库和神经网络知识库。对于问题领域比较清楚的可靠的知识,由领域专家向知识工程师提供,知识工程师负责将知识表示成规则放在知识库中;对于难以描述的知识,则用神经网络进行知识获取。在提供大量的实验数据和范例后,通过神经网络的学习,将数据和事例中的一般性规律存储在网络中。学习后的网络对新的实验数据和新的事例可以产生出符合要求的输出。

#### 4.3 说明

目前,国外已有基于工作站的智能CAD平台开始商品化。但在国内还没有形成面向设计的通用智能CAD支撑平台。而随着CAD技术在企业中的普及,面向设计的智能CAD支撑平台将有着十分广阔的应用空间。其体系结构也正是在这样的背景下产生的。但是,该体系是一个

十分庞杂的结构体系,在具体实施中将面临智能CAD领域中许多难点问题。若将其真正灵活、有效的应用到生产设计中,尚需付出艰辛的努力。

## 5 基于神经网络的CAD/CAM一体化

CAD/CAM辅助制造,是最具有代表性也是发展最快的现代制造技术之一,已经广泛应用于社会生产的各个领域,并成为衡量一个国家科技现代化和工业发展水平的重要标志之一。

目前单独的CAD/CAM软件已经比较成熟,但CAPP技术研究却相对落后,仍没有成熟的CAPP系统。航天工业中多采用国外的商业CAD/CAM系统,而国外的CAD/CAM系统的数据结构不公开,二次开发的能力有限,在此基础上创建CAPP系统是不可能的。所以必须对CAD/CAPP/CAM系统及其集成技术开展深入研究。

### 5.1 系统的结构

如图38.5-49所示,整个系统分为知识库、特征识别模块、智能设计模块、智能CAPP模块和CAM模块。知识库是整个系统的核心,存储零件和产品的几何、拓扑、加工、管理信息,存储用于控制各个模块运行的知识,担负各个模块的创造性工作和协调各个模块之间的信息交流与转换。知识库除了包含一般工程中所用的数据外,还包括生产实践中总结出的经验、理论和规律,包括参数选择、判定规则和决策优化等。

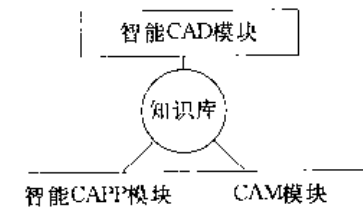


图 38.5-49 系统结构简图

知识库采用多种方法来描述生产中的知识:数据库、框架、产生式规则和人工神经网络。数据库用于存储设计、制造参数;框架则用来描述产品的结构、设计、加工信息;规则和人工神经网络则用于归纳生产中的创造性知识,并能根据已有知识进行归纳和推理。

智能CAD模块包括智能概念设计模块和机械设计模块,其中机械设计模块中包含结构设计和强度设计。结构设计是根据概念设计的结果——总体方案进行的具体结构设计,生成产品和零件的几何、拓扑信息,把他们存储在知识库中,作为CAPP/CAM系统的输入,并能输出图样。

通过自动设计模块或通过特征识别模块得到零件;产品和零件设计的数据,然后通过CAPP模块生

成零件加工的工艺文件,根据知识库中加工环境参数,CAM系统生成用于加工的NC程序。

### 5.2 产品零件数据结构

要解决CAD/CAM系统的一体化,首先必须有一个包含产品的设计信息、制造信息、装配信息和管理信息的统一数据结构。框架是专家系统一种重要的知识表达方式。框架是由节点和关系组成的一个局部网络。框架一般由框架名和槽组成,每一个槽都有其名称和对应的值,槽又分为若干侧面。槽代表事物的不同特性,而侧面是对槽的进一步说明,是对事物不同特性的细致的描述。槽或侧面的值可以是数字的或逻辑的,也可以是程序、条件、默认值或一个子框架。

零件框架包括四个槽:管理属性槽、设计属性槽、制造属性槽和装配属性槽。管理属性槽又包括多个侧面,分别表示零件的图号、名称、材料、工号等。设计属性槽的值是一个设计子框架,用于描述产品的设计信息。设计子框架由特征子框架组成。每一种特征子框架有各自独有的固定格式,如“圆柱面”特征共有四个侧面:特征编号、圆柱面的位置、圆柱面的尺寸(含直径、长度的尺寸和精度)、圆柱面的形位精度、表面粗糙度等。制造属性槽用于描述与零件制造有关的信息,如加工工序内容、机床、刀具、工艺参数等。装配属性槽则用来描述零件所属的部件或产品的名称以及与其他部件或零件的装配关系。

用框架表示零件信息的最大优点是可扩充性,可在设计和制造的任何阶段增加新的属性。由于系统采用一个整体的数据结构,在产品的设计、制造和管理的任何阶段都可以直接读取产品的所有信息。

### 5.3 智能CAPP系统

CAPP系统主要包含以下内容:加工链的选择、工序尺寸链的计算、工序内容的编排、工时计算、机床、刀具、夹具的选择等。在CAD/M系统中,工艺尺寸链计算采用了Hopfield网络,而加工链选择、机床、刀具、夹具选择采用BP网络来实现。以下就加工链的选择和工序尺寸链计算为例分别介绍。

#### 5.3.1 BP网络实现加工链的选择

加工链是特征的加工序列,是制造实现零件设计要求的保证,是CAPP设计的基础。加工链的选择考虑的因素很多,粗略概括起来,可以用函数表示如下:

$$P = f(M, G, D, Tol, S_j, Q, C_p, M_c) \quad (38.5-98)$$

式中  $P$  —— 所选择的加工方法;  
 $M$  —— 工件材料;

$G$  — 表面形状;

$D$  — 尺寸;

$Tol$  — 公差 (包含尺寸公差及各种形位公差要求);

$S_j$  — 表面粗糙度;

$Q$  — 生产批量;

$C_p$  — 生产费用;

$M_c$  — 可使用的机床设备。

加工链选择实际上就是根据实际情况进行匹配的过程。

在传统的创成式 CAPP 系统中, 一般采用判定表 (或判定树)、产生式规则来表示。由于在实际生产中要考虑的因素很多, 用判定表、规则表表示时层次多, 知识的整理也十分复杂, 而且很难囊括所有可能的情况, 所以, 系统只能检索存储的知识, 没有容错能力, 系统的鲁棒性差。用 BP 网络处理这类非线性映射却能消除这些不足。根据航天惯性器件零件的加工特点, 我们总结了 12 项指标, 作为 BP 网络的输入矢量, 输出矢量为 4 项, 隐层节点数目为 10, 4 位输出矢量分别表示粗加工、半精加工、精加工、超精加工。

输入矢量的含义如下:

1~3 位表示加工特征的类别: 车加工、钻镗加工、铣加工。其后分别代表关键尺寸大小、尺寸精度、形状参数、粗糙度、形状公差、位置公差、材料的加工性能、零件的重要度系数和调整系数等。

BP 网络中输入矢量和输出矢量都采用 [0, 1] 间的单精度数表示, 这就要求把从生产中搜集到的知识进行量化和规范化处理。由于特性具有相似性和模糊性, 用模糊数学的隶属函数来进行量化和规范化处理。以尺寸特性为例。

根据公差理论, 公差等级与公差数值有以下关系:

$$IT_1 = 0.008 + 0.02 \times D \quad (38.5-99)$$

$$IT_5 = 0.45 \times \sqrt[3]{D} + 0.01 \times D$$

在  $IT_2 \sim IT_5$  之间, 有

$$IT_n = IT_1 \times \left( \frac{IT_5}{IT_1} \right)^{n-1} \quad (38.5-100)$$

$IT_5$  以后, 每增加五级, 公差增大 10 倍。

由于在实际加工中,  $IT_1 \sim IT_3$ ,  $IT_9$  以上, 加工特性变化不大, 而在 5~8 级之间变动较大, 所以采用降半岭型函数:

$$f(x) = 0.5 - 0.5 \sin \left[ \frac{\pi}{6} (x - 6) \right] \quad (38.5-101)$$

得到的分布如表 38.5-8 所示。

表 38.5-8 尺寸隶属度

3	4	5	6	7	8	9
1	0.933	0.75	0.5	0.75	0.07	0

由此可以看出, 3~9 级之间, 其函数值的离散性较大, 反映到网络中就是不同样本间的距离较大, 从而达到较大的识别精度。

其他矢量位用类似的方法处理, 把已有的成熟工艺量化、规范化处理后输入样本, 网络就可以根据样本进行训练了。具体算法见参考文献 [3]。

### 5.3.2 工艺尺寸链计算的 Hopfield 网络

目前工艺尺寸链的传统方法是采用工艺尺寸链图解算法, 但这种算法用计算机实现起来比较难, 而且容易出错, 效率不高。尺寸链的计算实际上是一个路径优化问题, 可以用 Hopfield 网络来求解。现以回转体零件轴向尺寸为例。

网络的节点表示两个端面之间有无已知的尺寸和公差。如果  $i$  端面和  $j$  端面间的尺寸已知, 则  $x_{ij} = 1$ , 否则为 0。节点之间的权值用  $W_{i,mn}$  来表示, 选值原则如下:

1) 节点自身无反馈, 即  $W_{i,mn} = 0$ 。

2) 如果  $i \neq j$

$$W_{i,mn} = \begin{cases} 1 & m=n=i \text{ 或 } m=n=j \\ 0 & \text{其他} \end{cases} \quad (38.5-102)$$

3) 如果  $i = j$

$$W_{i,mn} = \begin{cases} 1 & m=i \text{ 或 } n=i, m \neq n \\ 0 & \text{其他} \end{cases} \quad (38.5-103)$$

节点的特性函数表示为

$$f(x_{ij}) = \text{sgn}(x_{ij}) \quad (38.5-104)$$

神经元阈值为

$$\theta_{ij} = \begin{cases} 2 & i=j \\ 1 & i \neq j \end{cases} \quad (38.5-105)$$

系统的能量函数为

$$E = \sum_{i,j=1}^M f(x_{ij}) \quad (38.5-106)$$

网络的运行方程为

$$x_{ij}(t+1) = f \left( \sum_{m,n=1}^N W_{i,mn} x(t) - \theta_{ij} \right) \quad (38.5-107)$$

算法如下:

① 网络节点权值初始化。

② 若第  $i$  端面和第  $j$  端面之间的值已知, 则令  $x_{ij} = 1$ , 否则为 0。

③ 若要求端面  $m, n$  之间的值, 如果  $x_{mn} = 1$ , 则说明值已知, 输出值并停止计算, 否则使  $x_{mn} = 1$ , 进行下一步。

④ 计算网络的能量值, 如果能量值达到最小, 则

结束, 否则进行⑤。

⑤ 依次计算各个节点的值, 重复①。

最后, 节点为1的各个节点就是待求尺寸的封闭环。从待求尺寸出发, 依次把各个节点连起来, 节点号增大的为增环, 节点号减小的为减环。这样就可以求出任意的尺寸值与公差, 也可以反求加工尺寸的大小和公差。

Hopfield 网络求解尺寸链, 不受节点个数、尺寸标注复杂性影响, 具有自组织、自学习的特点。网络取决于权值、节点函数、能量函数。不同的 Hopfield 网络的计算过程基本相同, 所以计算时可以用同一个数组, 节省了系统的内存, 而且如果用硬件实现, 可采用并行计算, 其速度是其他方法所无法比拟的。

#### 5.4 CAM 模块

CAM 系统的功能主要是把 CAPP 产生的工艺文

件与实际生产环境(机床、刀具、夹具)相结合, 生成适于生产的 NC 程序。这个过程主要包含建立刀具参数文件和刀具轨迹文件, 生成刀位文件, 并能根据实际加工机床转换成 NC 程序, 经仿真检验后输送到前端机。

系统所建机床数据库的格式见表 38.5-9, 采用 Access 数据库进行存储。刀具和夹具数据库的格式与之基本相同。

其中切削参数对应的是一个切削参数表。

人工智能技术可以解决设计和制造中的知识输入的“瓶颈”, 从而把生产实践中总结出来的知识应用到计算机辅助设计和制造中, 特别是人工智能技术的引用, 解决了 CAPP 系统的知识获取和处理的问题。一体化系统必须建立完整的可扩充的数据格式, 并在此基础上实现一体化。

表 38.5-9 机床数据库

机床编号	机床类型	机床自由座	可加工范围	机床状态	数控系统	加工精度	刀具库容量	循环类型	差补类型	辅助功能	切削参数
CH01	车	2.5	100×200	好	S	0.001	8	有	直线	1	3
XT03	铣	3	350×300	好	F	0.002	24	有	圆弧	2	5

## 第6章 人工生命设计技术与方法

### 1 人工生命技术基础

#### 1.1 人工生命的进化模型

人工生命研究的重要内容之一就是进化现象,而遗传算法则是研究进化现象的重要方法之一。遗传算法的基本内容在前面各章已有详细叙述,它采用符号序列来描述信息集合,然后通过一些遗传操作,如交叉(即符号序列的混合)、突然变异(生成符号序列的新规则)、选择(选取最优符号序列)、淘汰(去除剩余符号序列)等,得到一些优化解。进一步,可以把上述遗传操作反复执行,以得到最优解。若把它与能够分析生命的个体或集团行为的博弈理论结合起来,则可进一步提高人工生命对生态系统的适应性。因此,遗传算法是人工生命研究的重要理论基础之一。

在这一节中,着重讨论人工生命的生成与进化模型和遗传算法关系比较密切的几个问题。为此先来看看人工生命的生成结构。

表 38.6-1 人工生命的生成结构

生成结构	生成方法	例
生物体内部系统	建模法	神经网络、免疫网络、元胞自动机、L 系统等
	动作原理法	基于混沌、分形、元胞自动机等组织化
生物体外部系统(实体集团系统)	建模法	遗传算法、博弈理论等
	动作原理法	伴有自组织化的分布式协调原理

人工生命的生成结构如表 38.6-1 所示。它主要分为两大类:一类是构成生物体的内部系统,主要包括生物体中的大脑、神经系统、内分泌系统、免疫系统、遗传系统、酵素系统、代谢系统等;另一类是生物实体及其集团所表现的外部系统,主要包含生物实体集团对环境的适应系统和遗传进化系统等。因此,可以从生物内部和外部系统来获取各种各样信息,用这些信息生成人工生命。就其生成方法来说也分为两种:一种是建模法,即先把由内部或外部系统获得的生命行为信息模型化,然后再由这些模型生成生命特有行为;另一种是动作原理法,即基于混沌、分形等原理的生成方法。混沌、分形原理可以用来描述生命行为的原理,因此生命行为是自律分布的非线性行为。

例如,直接应用现代计算机技术的人工神经网络系统就是属于生物内部系统范畴的建模系统,而遗传

算法则是属于生物外部系统范畴的建模系统。事实上,在神经网络信息处理中,其处理行为就包含着混沌现象。遗传算法可以被认为是自律分布的并行处理方法。因此人工生命的产生就是从这样一些模型系统所表现出的各种各样生命固有行为出发,把它们的行为原理概括为一些基本算法,用这些算法来生成人工生命。

可见,遗传算法可以用来研究生物体外部系统(也就是实体集团系统)中生命行为规律,而这种规律往往表现出自律分布的并行特性。以下简述人工生命的进化模型与遗传算法有着许多共同特点的问题。

##### (1) 个体表现问题

即使是表现相同行为时,某个体如何表现,要决定于该个体所属搜索空间的结构和大小。这种搜索空间的结构,决定了所谓的“适应度地形”(即淘汰值曲面,该曲面与搜索策略二者决定了进化能力),也就是搜索空间和搜索策略决定了人工生命的进化能力。

例如,Joshua R. Smith 从昆虫的进化角度,给出了以下的表达。其定义昆虫集团的染色体为 16 个。每个昆虫的染色体用  $2n$  个基因表达,其图像在  $X$ 、 $Y$  二维空间的坐标值用傅里叶系数  $(A_1, \dots, A_n, B_1, \dots, B_n)$  表达 (Joshua R. Smith, 1990, 1991):

$$X = \sum_{i=1}^n A_i \cos(it) \quad (38.6-1)$$

$$Y = \sum_{i=1}^n B_i \sin(it) \quad (38.6-2)$$

在  $n=8$  条件下,随着  $t$  的增加可以描绘出昆虫的形态。在遗传进化过程中,通过对基因型的增殖、交叉、突然变异,其进化如图 38.6-1~图 38.6-4 所示发展下去。

总之,通过个体表达使人工生命的传感器/效果器具有可变性,使个体的行为、形态受到较小的限制,是人工生命都面临的一个难题。

##### (2) 搜索策略问题

在进化搜索空间内,如何设定搜索点的“转移规划”,是人工生命搜索策略的一个重要问题。在 GA 中,有“淘汰·增殖”、“交叉”、“突然变异”、“反馈”等遗传操作,这些遗传操作的实用形态(或方式)直接影响搜索能力。在人工生命中,与个体的表达相关联,其基准仍然是,搜索策略的制定,决定其操作能力与淘汰值曲面形状的组合,是否能使系统脱离局部解,即决定了它是否具有进化能力。

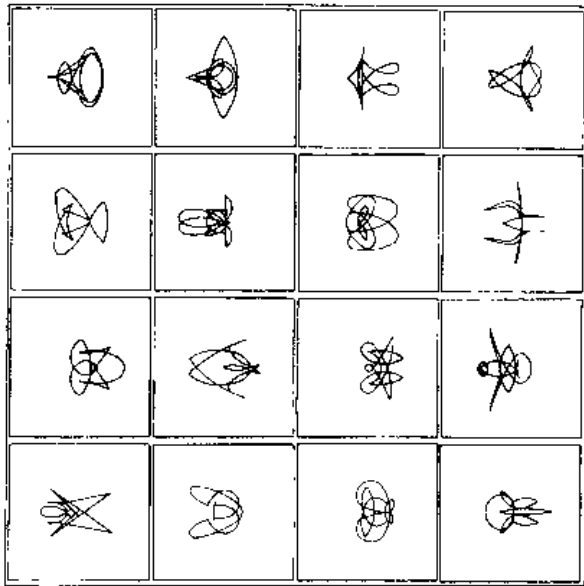


图 38.6-1 昆虫的第一代进化形态

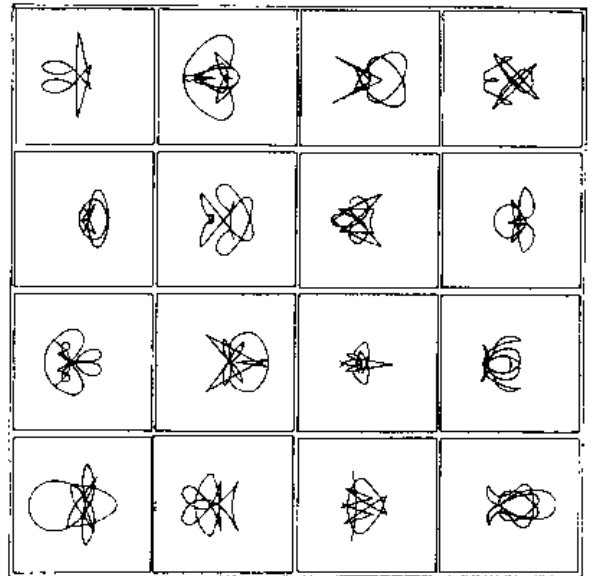


图 38.6-3 昆虫的第三代进化形态

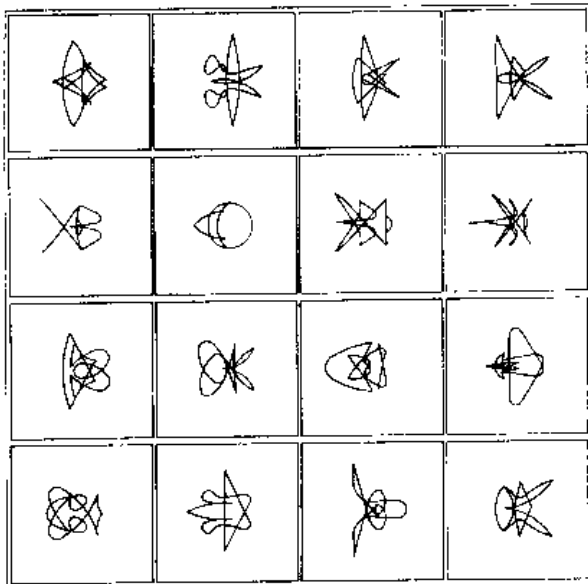


图 38.6-2 昆虫的第二代进化形态

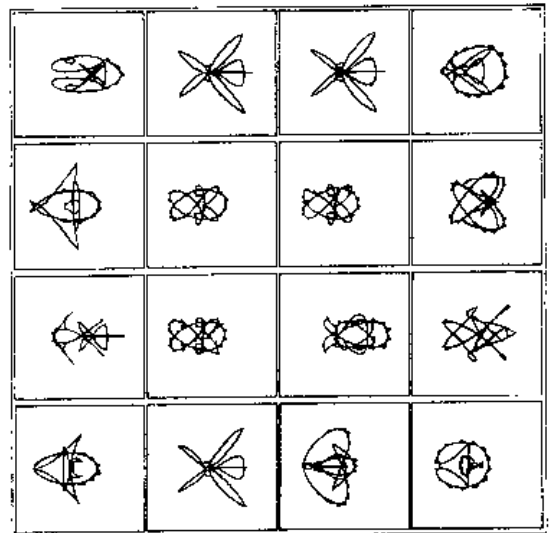


图 38.6-4 昆虫的第四代进化形态

仍然以昆虫捕食系统来说明。在昆虫捕食的搜索空间内，如图 38.6-5a 所示，昆虫总是向食物（细菌）浓度高的地方行进。

昆虫的基因由图 38.6-6 所示的  $F, R, IIR, RV, HL, L$  移动方向表达。假如  $(F, R, HR, RV, HL, L) = (2, 1, 1, 1, 3, 2)$ ，昆虫将可能选择向对应较大的数值方向移动。昆虫在移动过程中，如图 38.6-5 所示，其年龄、能量、基因码随之发生变化，即在移动过程中成熟、生殖、死亡。由于这些变化，假如对于某些昆虫，其染色体中  $(F, R, HR, RV, HL, L)$  的移动方向值没有明显的较大值，可能会出现如图 38.6-7a 所示的移动方向犹豫虫。犹豫虫将自己周围的细菌食完后，因不能广域搜索最终饿死。如图 38.6-

7b 所示的直行虫，在集团中具有优势，可得到充分的进化。典型的直行虫（例如，其染色体为  $(F, R, HR, RV, HL, L) = (9, 6, 0, 2, 4, 1)$ ），其基因具有以下特征：

- 1) 向前 ( $F$ ) 移动的基因值大 ( $F=9$ )。
- 2) 向后退 ( $RV$ ) 的基因值小 ( $RV=2$ )。
- 3) 向右 ( $R$ )、左 ( $L$ )、右后 ( $HR$ )、左后 ( $IIL$ ) 的基因值不大不小 ( $R=6$ )。

第 2) 个特征是极其重要的。因为持有大  $RV$  基因的虫，易于出现回转现象（如图 38.6-7c）的回转虫，其极易死去。其次，虫到达搜索边界，如果不能反方向行进，也易于饿死。为此第 3) 个特征可能使虫具有回转捕食的聪明行为。

假如以山的高度对应细菌浓度的表达，那么，此昆



虫捕食系统将成为一个最优化问题。考虑图 38.6-5 的虫的信息、有性生殖和无性生殖等,描述昆虫染色体可用以下三个参数来表达:

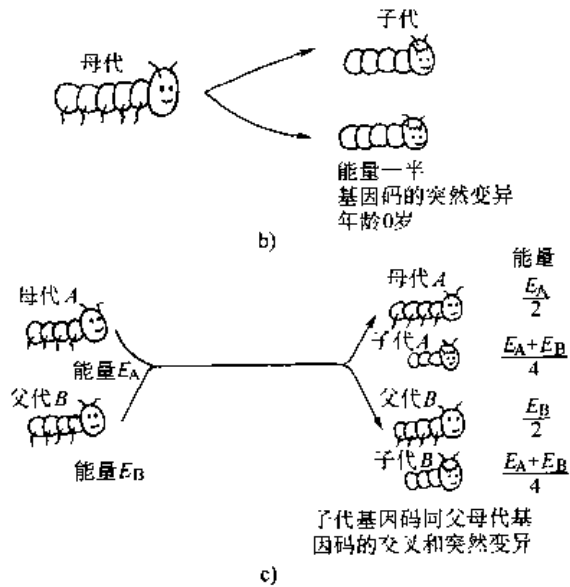
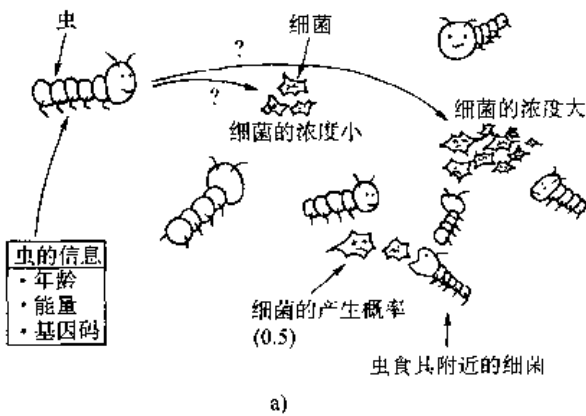


图 38.6-5 捕食昆虫的搜索与表达

a) 食细菌的虫群 b) 无性生殖 c) 有性生殖

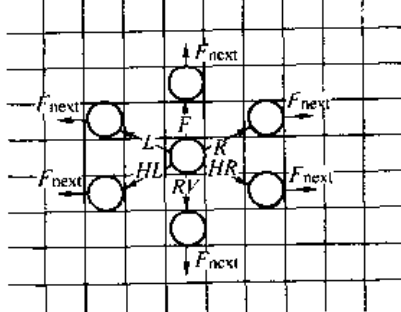


图 38.6-6 捕食昆虫的移动方向

位置:  $X(t) = (x_1(t), \dots, x_n(t))$

方向:  $\vec{DX}(t) = (dx_1(t), \dots, dx_n(t))$

能量:  $e_i(t)$

昆虫登山的搜索问题可以形象化表达为图 38.6-8。其进化过程用图 38.6-9~图 38.6-11 简单表述。可见,人工生命的进化搜索策略,除恰当地利用遗

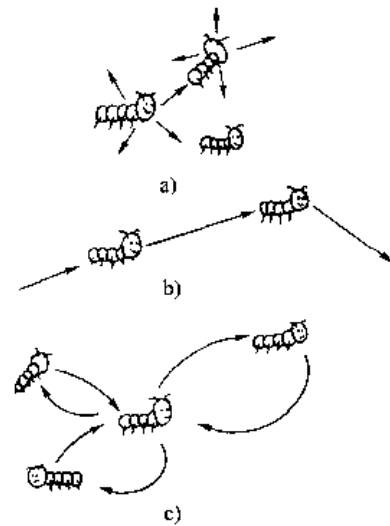


图 38.6-7 昆虫的移动方向性  
a) 犹豫虫 b) 直行虫 c) 回转头

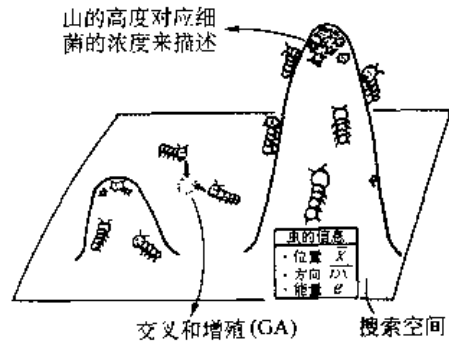


图 38.6-8 昆虫登山搜索

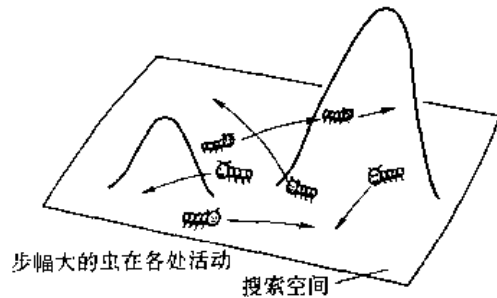


图 38.6-9 GA 搜索的初始阶段

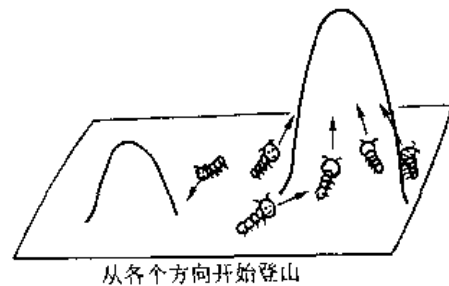


图 38.6-10 GA 搜索的中期阶段  
传操作外,需要凭借操作者的经验和直觉“自行搜索”,或者引入新的操作,提高搜索效率。



这样不仅植物可以生长得高大茂盛,而且在各次迭代阶段也有相同的复杂度和相似度,需要强调的是,替换只是针对串中的变量 ( $F$ ) 进行的,而对标识符 (+, -, [, ]) 则不作任何替换,直接保留,一般把

【基本结构, 分枝类型, 单位长度, 单位角度】

(38-6-6)

称作该植物的基因,它决定了一个特定植物的概念形态,与最终的图形比起来,图 38-6-12a, b, c 分别为由基因

【“1”,  $F$  [+ $F$ ] [- $F$ ], 1, 36°】(38-6-7)

分裂繁殖 1 代、2 代、3 代的结果,其具有数据量很小的特点。综上所述,可以得到一个由  $L$  系统产生植物与树的流程图,如图 38-6-14 所示,这里,没有考虑树的花、果实、叶子等,也没有引入任何植物学中的名词,它确实很简单,但是这种模型对植物的形态生成,以及在工程上的模型表达方法上具有重要的意义。

可以肯定,如果将  $L$  系统的各种参数作为基因进行进化,将得到更为复杂的仿真形态。其基本操作与 Smith 模型相同,可以得到如图 38-6-15 的“生物形态”

态” (Dowkins, 1989 年)。

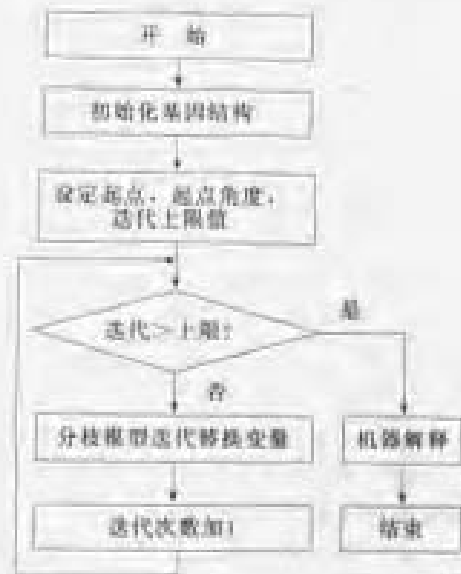


图 38-6-14 L 系统模拟植物与树的流程图



图 38-6-15 各种各样的“生物形态”

## 2 人工生命的研究内容归纳<sup>[40]</sup>

从近年来关于人工生命的研究内容大致可以归纳为以下几个方面:

### 2.1 数字生命的研究

所谓数字生命专指以计算机为工具和媒体,计算

机程序为生命个体的人工生命研究,这方面以 Thomas S. Ray 的数字生命世界 Tierra 为代表, Thomas S. Ray 是一位生物学家、进化论学者,他把生物学上有机体进化的概念引入计算机领域,用数字计算机所提供的资源为他的数字生命提供一个生存环境,他设计的数字生命以数字为载体,探索进化过程中所出现的各种现象、规律以及复杂系统的突现行为,数

字生命利用 CPU 时间来组织其在存储单元中的行为。数字生命以一定的计算机程序形式存在于 RAM 环境中,它为占据 CPU 运行时间、存储空间而通过响应的竞争策略相互竞争。一个“生命”必须被设计为适合在这样环境中生存的某种数字代码程序。这个程序能够自我复制,并且直接被 CPU 运行。这些机器代码能够直接触发 CPU 的指令系统以及操作系统的服务程序,通过对资源的占有来体现它在进化过程中的优势地位。

在 Tierra 的运行中,随着世代的推移,生命体呈现出复杂的现象,种类日益增多,同时“单细胞”向多细胞进化,形成自己的生态环境。在生命的进化过程中,曾经出现过物种大爆炸的情况。如今, Tierra 运行于全球 150 个网络环境之中,其复杂程度还在不断增加。

数字生命的研究中一个重要的模型就是元胞自动机,元胞自动机被认为具有突现计算(emergent computation)功能。由于人工生命研究的重要内容是进化现象,遗传算法是研究进化现象的重要方法之一。

## 2.2 数字社会的研究

Joshua M. Epstein 和 Robert Axelrod 在计算机上创立了一个数字社会 Sugarscape。这个人工社会用来研究文化和经济的进化过程。他们认为一个人工社会是这样的计算机模型,它包含一群具有自治能力的行为者;一个独立的环境;管理行为者之间、行为者与环境之间以及环境各个不同要素之间相互作用的规则。人工社会的行为者是一个能够随着时间发生变化或者具有适应性的数据结构。每个行为者具有遗传特性、文化特性,以及管理它与环境和其他行为者之间的规则。其中行为者的遗传特性在其生命周期内是固定的。在 Sugarscape 中行为者的性别、新陈代谢以及视野是其遗传特性,而文化特性是由父母传给子女,并通过与其他行为者的联系而横向地发生改变。其环境里含有可更新的能源——糖,行为者依赖糖组织自己的新陈代谢。人工社会是由各个行为者自我组织形成的,由各个行为者在简单规则的支配下,与人工环境交互作用突现形成的。

## 2.3 虚拟生态环境

挪威的 Keith Downing 提出了名为 EUZONE 的一个进化的水中虚拟生态环境,目的是提供一个观察生态系统是如何从原始状态进化以及复杂生态系统突现行为的实验手段。它利用具体的物理和化学模型,结合进化规划建构以碳元素为基础的水中生态环境,可以观察到低等动物形体的进化及生存竞争。EUZONE

具有两个基本过程:环境模拟和生物的进化。环境模拟尽量反映真实世界的物理、化学以及生物之间的相互作用,生物进化由遗传程序设计和遗传算法来实现。

## 2.4 人工脑(Artificial Brain)

日本的 ATR 的进化系统部(evolutionary system department)致力于开发新的信息处理系统,这种系统具有自治能力和创造性,他们将这样的系统称为“人工脑”。人工脑不仅能自发地形成新的功能,而且能够自主地形成自身的结构。其研制者并不想单纯地再现生物大脑的功能和结构,而是要得到在某些方面优于生物大脑的信息处理系统。

人工脑采取两方面的实现方式:类似生命的模型(life-like modeling)和社会模型(social modeling),包括传统的用于神经系统的学习模型(如人工神经网络)。在类似生命的模型中,系统有一个类似于生命系统胚胎发育的功能,使得系统的结构和组成单元能够发生变化,形成复杂系统。在社会系统中系统被视为动态过程,在这个过程中,局部的、各个单元之间的连接使得整体的、全局的功能及次序、状态发生突现。反过来,各个单元也受到全局状态的影响。因此两个方向的相互连接、影响系统发生变化。为使系统具备自治和创造性,系统本身需要一些机制在功能和结构上的自发变化。研制者在系统设计中引入“进化和突现”的极值。

ATR 对于人工脑的建构正在从硬件和软件设计两方面来推进。这个项目在数字计算机上通过自然选择的简化来产生复杂和智能化的软件。进化的基本因素是带有可遗传变异的自我繁殖。为了在硬件上实现进化计算,需要一种特殊的硬件平台,可进化的硬件目前处于开发初期阶段。大规模的神经网络和极高速要求,需要高容量的存储器以及高速的电子器件,CAM Brain 项目运用“进化工程”(evolutionary engineering)技术来建构、发展、进化出以 RAM 和元胞自动机为基础的人工脑。

## 2.5 进化机器人(evolutionary robotics)

生物系统给人们提供了分布式控制的思路,其脑神经系统、遗传系统、免疫系统的功能启发了人们把生物学上的一些现象工程化,并且运用于机器人的设计上。目前正在发展的第三代机器人要求具有人的简单智力和学习能力。

Rodney A. Brooks 提出了基于行为的设计方法。此方法在 20 世纪 80 年代中期开始使用,设计出比传统设计方法行动更快和更灵活的机器人。进化机器人的操作方式是自律型的,其位置、移动等是突现形成

的,其智能也是由各个并行执行的小过程自组织突现形成的,并且这样的小过程分散在整个系统中。进化机器人具有比传统机器人更快的速度和更好的灵活性、鲁棒性,进化计算可以比较容易地植入到这样的系统中,其硬件和软件的设计以及测试费用比以前要少。

## 2.6 进化软件代理 (evolvable multiagent)

人工智能的研究人员长期以来一直在研究一种复杂的建构软件代理的方法。例如,一个具有人工智能的电子邮件 Agent 可能知道有行政助理人员,知道某位用户有一名叫 George 的助理,知道助理必须掌握老板的会议日程,还知道“会议”这个词的信息可能含有日程信息。有了这些知识,这个 Agent 就可以推导出它应当转送此信息的复制件。

以知识为基础的软件代理要求包括所有常识信息的知识库,但一般软件工程师只能系统地整理比较狭窄领域的知识。采用人工生命的方法进化软件 Agent 可能是最有发展前途的方法。“人工进化”可以随着时间的推移整理出一个系统中的最有效的代理人(由其主人评定)的行为,并把这些行为结合起来以培养出适应能力更强的群体。可以设计这样的电子邮件 Agent,它们能够连续观察一个人的行动,并把他们所发现的任何有规律的行为实现自动化。电子邮件代理观察到用户总是把含有“会议”的信息的复制件转交给行政助理,由此便可领悟到其规律,然后自动做这项工作,此外,Agent 可以向执行同一任务的 Agent 学习,例如,一个电子邮件代理在遇到一份陌生的信件时可以询问它的同伴,从而得知人们通常是看了私人递交给他们的电子邮件后,再看按邮送名单递交的电子邮件,这类合作可以使一群代理以复杂的、明显智能的方式行动。

将人工生命 Agent 置于新一代计算机网络中,形成一个电子生态系统。对用户有用的或对其他 Agent 有用的 Agent 将运行比较频繁,从而得以生存下来并繁殖后代,那些用处不大的 Agent 将被清除掉。随着时间的推移,这些数字化生命形式将占据不同的生态环境。有的 Agent 可能进化成优秀的数据库编制者,其他的 Agent 则是使用它们的索引来找到某一用户感兴趣的文章。可能会出现寄生、共生、免疫以及生物世界中常见的其他现象在计算机网络中的实例。随着外部对信息的要求发生变化,这个软件生态系统将连续地更新自身。

IBM 公司目前正在开发一种被称为电脑空间免疫系统的软件。正如脊椎动物的免疫系统在一种新病原侵入机体后几天之内就会产生出对付它的免疫细胞一样,计算机免疫系统可在几分钟内就能产生出识别

并消除新遇到的计算机病毒的方法。

## 3 人工生命的设计方法

人工生命为解决问题提供了新的思想与工具,其研究开发有重大的科学意义和广泛的应用价值。人工生命的研究与开发有助于创作、研制、设计和制造新的工程技术系统,如人工脑、智能机器人、计算机动画的新方法。数字生命、软件生命、虚拟生物可为自然生命活动机理和进化规律的研究探索提供更高效、更灵活的软件模型和先进的计算机网络支持环境。利用人工生命研究人类的遗传、繁殖、进化、优选的机理和方法,有助于人类的计划生育、优生优育;利用人工生命研究动物的遗传变异、杂交进化的机理和方法,用于发展动物的新品种、新种群;利用人工生命研究植物的生长、杂交、嫁接、移植的机理和方法,用于发展植物的新品种、新种群。人工生命的研究开发及应用将进一步激发和促进生命科学、信息科学、系统科学等学科的更深层的、更广泛的交流和新的发展。

### 3.1 金融证券市场分析决策中的人工生命应用

国际上目前运用人工生命算法进行金融证券市场分析决策研究处于蓬勃发展的阶段,有效的运用拟生态技术进行复杂金融市场分析是一个大课题。

人脑自生命诞生以来,经过数十亿年漫长岁月的进化,形成了具有高度智能的复杂系统。人脑不必采用复杂的逻辑运算,却能灵活处理各种复杂的、不精确的和模糊的信息,善于理解、发现、创新、决策和具有直觉感知等功能。大脑在结构上是由 140 亿个神经细胞组成的大规模网络,他的各种智能功能都是这个大规模网络处理的结果。人工生命算法以大脑细胞的生理机能作为研究对象,运用计算机对大脑的基本单元——神经元进行数学模拟,建立的一种由神经元组成的“大脑模型”,这就是人工生命算法雏形。

在证券市场里,千千万万个持有资金大小不等、投资理念各异、投资方法多样的投资者面对一千多只股票分别作出各自的决策,再加上各种政经信息、上市公司经营表现、市场传闻等等因素的影响作用,导致了金融证券市场是一个内部规律极端复杂、较难预测把握的大系统。要想在一个复杂的、每天都推陈出新的系统当中立于不败之地,只有用手中的方法、工具很好的表述这个市场,真正从本质上抓住运作规律,实现所谓的无招胜有招。运用简单的统计运算方法对市场的基本数据进行处理,从而希望揭开蒙在市场规律上的面纱,哪怕是揭开一个小角找到一点点规律就是成功的。比较典型的有均线系统、各种技术经济分析指标、猜想性

筹码分析理论等。

光揭开市场规律的一角是远远不够的,因为这个被揭开的角是片面的,是局部规律,它往往会让你做出错误的决策,因为它不是全局把握的规律。人们一直有一个梦想,就是希望拥有叫“当猫”的知识面包,只要把知识面包放在书上,书上的知识就印在面包上,吃下这个知识面包你就学会了书上的知识,多省力啊。我们就是在打造这样一个知识面包型的人工生命体,因为要解开市场运作的规律光靠人有限的精力是不行的,就算一个非常成功的职业投资者,几十年的投资经历所积攒的经验可能就是那么一两条,光靠这么一两条经验或投资方法就可以很成功了,但代价就是几十年如一日的在失败中学习教训、在成功中积累经验。我们的人工生命算法造就的活性生命体不仅能给它的知识、规律学会,更重要的是它能不知疲倦的主动去寻找没有学过的规律,就像一个不仅会印知识,还会到处主动找书来印的知识面包。这样我们的生命体就能不知疲倦的在海量的股市数据中不断的发掘其中的运作规律,成为一个靠增长知识、规律来生长的人工生命活体。

人工生命算法用于证券分析有以下优势:

1) 自学习、具有生长能力 人工生命体不仅能把目前绝大部分有限的操盘手法、定势完全学会,同时能每天不知疲倦的在海量的股市数据中不断的发掘其中的运作规律,不断发现各种变化的手法和规律,成为一个靠增长知识、规律来生长的人工生命活体,如当政经等大环境及上市公司内部环境因素发生变化时,不少庄家会改变其操作手法以蒙骗股民,人工生命体经过对历史上所有股票相关信息的人工生命矩阵的训练、学习,达到紧跟潮流,永不落伍的目的。好似冲浪,长在浪尖。

2) 自适应及推广能力强 人工生命体学习生成后,由于人工生命的抗干扰性和自学习能力,在一只股票中形成的生命矩阵能够用于其他股票。比如同一庄家在不同股票的操盘手法进行发现和跟踪,即使只有几只股票的信息都能够用于分析预测其他股票的走势,甚至用于推测大盘的变化趋势,所谓“落叶知秋”。同理,即使只有某一时段的股价信息,也能够预测分析出走势规律。

3) 容错能力强 人工生命有人脑的多细胞容错能力强的特点,在追踪同一种操盘手法规律时,如果庄家有意采取短期的打压或强行拉升股价等手法,以迷惑其他投资者时,人工生命体能有有效的识别这种某阶段的操盘手法的变异。

4) 抗干扰能力强 人工生命能够全面分辨规律的细节,能从本质上把握操盘手法的规律,从而使庄家

表现在价量关系上的干扰性操作手法被区分出去,对有意拉长、缩短操盘手法周期或少量减少部分操盘进程等手段能有效包容,达到抗干扰的目的。

5) 可并行计算、速度快、可处理海量数据 人工生命的特点使之能够同时处理大量数据,人脑比电脑慢得多,但在识别复杂事物的能力方面却比当今最快的电脑还快。现在不同了,人工生命矩阵具备了人脑这方面的优点,同时克服了人脑速度慢、容易疲劳的缺点,对海量数据并行处理分析,得到令人满意的答案。

6) 宏观把握能力强、微观探索时细致、周密、信号量大 人工生命算法能够解决传统算法的单一性和简单性,能够从全局的角度把握股票市场的规律,达到规律自动发掘和规律全面发现的目标,能够同时发现同一或类似操盘手法在不同股票和不同市场的表现规律,这样既能够把握市场的重大机遇,也不放过细小机会。

### 3.2 计算机动画的人工生命应用

生活在自然界的动物群体是动画创作者们面临的“挑战性”难题。而“人工生命”方法,可以逼真地体现自然生态系统中动物群体的复杂运动和行为,并且可显著地降低动画创作者的劳动强度。

其基本方法是构建“人工动物”,创作自激励的自主智能体,去模拟动物个体的真实外观、运动和行为,以及动物群体的社会行为表现模式。用计算机模型描述这些人工动物共有的基本特征——生物力学、运动、感知和行为。在虚拟海洋世界栖息着的各式各样的人工鱼群证实了“人工生命”的有效性。每一条人工鱼就是一个自主智能体,它有基于物理的、可变形的、由内部肌肉驱动的肌体;有感觉器官,例如眼睛;有感知、运动和行为控制中枢的鱼脑。每条人工鱼通过肌肉运动的协调控制,可在虚拟的水流中游动。这些人工鱼展现出一系列的自然行为:在栖息地寻觅食物,绕过障碍物游行,与捕食者斗争,纵情于求爱仪式以获得配偶。类似于自然鱼群,人工鱼群的行为也是基于它们对外部动态环境的感知和内部动机和习性的。

由于人工鱼的行为能自动适应虚拟的水中环境,它们的运动细节无须动画师的详细刻画或规定。用传统动画制作方法,动画中的动物只是没有任何自主性的三维几何图形,好像没有生命的木偶,动画师在动画制作过程中的角色类似木偶戏的表演者。在计算机图形学中,大多数动物动画的制作是采用传统的、花费大量劳动的“关键帧”方法,计算机只是用来制作“关键帧”之间的中间帧。相反地,采用“人工生命”的动画制作方法,动画中的人工动物是自激励的自主智能体,

好像是有生命的真实动物，动画师扮演的是动物世界电影摄影师的角色。通过建立人工动物和它们的生存环境模型来生成动画，将自然生态系统的动画生成看作是动物在栖息地生活的可视化仿真过程已经跨越了“计算机图形学”和“人工生命”两个领域。然而，要将自然生态系统在屏幕上表现得和真实世界一样逼真和迷人是很困难的，主要原因是它本身的复杂性。在一个动画系统中，可能会有许多动物，每个动物都表现出不同的行为。理想情况下，动画师希望以最少的劳动获得丰富的自然景观。如果不仅要表现栩栩如生的动物形态，而且要表现每个动物的运动和它们的行为，其困难程度可想而知。生态系统是由动物和它们生活的环境组成的，动物的行为和它们生存的动态环境是息息相关的，尤其是和其他动物之间的关系。因此，人们在评价动画系统的效果时，有非常严格的标准，即使是——点点小的缺陷，都很容易被看出来。逼真的视觉效果并不是这类动画的唯一要求，要想用在娱乐和教育领域，动画师还应能控制动画的各个方面，尤其是要能够方便地修改动画。例如，改变虚拟环境，变换人工动物的数目、种类及分布的位置，改变动物的性格与它们的相互关系。

传统的计算机动画方法，如“关键帧”方法，制作了不少出色的动画，包括一些动物的动画。但是，它存在一些问题：

#### (1) 动画角色缺乏自主性

由“关键帧”方法制作的动物角色缺乏自主性，因此降低了动画系统的灵活性和交互能力。由于动物的真实行为受它所处环境状态（如不可动的和可动的物体、树木、岩石和其他动物）的影响，轻微地修改动画剧本，如移动一棵树或添加另一个动物，都需要将整幅动画重新制作，图形中的角色是不能与它所处的周围环境自动协调的，因此，在制作虚拟现实、计算机游戏和交互式教学工具时，采用“关键帧”方法是很困难的。

#### (2) 自然真实性难以保证

采用传统的“关键帧”方法，动画的真实性只能依赖动画师的技巧，除非动画师有很高的技艺水平，否则可能会出现糟糕的视觉效果，特别是由于通常的几何模型不具备力学特性，因此，动画中动物运动的物理正确性是没有保证的，动画中的角色也不会真实地响应外力的作用，动物之间及其周围环境的关系显得很协调，缺乏自然真实感。

#### (3) 低水平的动作细节控制

采用传统的动画方法，动画角色的每一个动作都要由动画师一一规划，也就是说，动画师要完全控制动画的每一个细节，例如制作卡通片。但是，许多应用系统对这种低水平的细节控制并不感兴趣，例如创作一

个虚拟现实的动物园：某个动物是否在某个时间、以某种姿态出现是不重要的，重要的是狮子和猴子要看起来像真的，它们的运动和行为要逼真。在这种情况下，我们并不希望动画师花费大量劳动去控制动画的每一个细节，因为这意味着动画中的角色只有很少的或根本没有自主性。因此，应当放弃这种低水平的动画控制，去寻求一种高水平的动画控制的方法。

#### (4) 需要动画师的大量劳动

传统动画方法最为成功的应用是在一部很卖座的影片“侏罗纪公园”中制作了恐龙。这些恐龙尽管看起来像真的，但它们只是些图片。它们的每个动作和移动细节，都是有高超技巧的动画师们一步一步设计的。这也就显出了“关键帧”方法的缺陷：随着动画的加长、复杂程度和真实性要求的提高，动画师们的劳动量将显著地增加。

#### 计算机动画的“人工动物”方法的优点：

1) 动画角色的外观、形态、运动和行为在视觉上令人信服。

2) 动画角色具有很高的自主性，不必花费动画师的大量劳动进行干预，可以自行完成动画的生成过程。

3) 动画角色应接受高水平的控制。动画师可以在高水平上控制或指挥动画角色的行为，例如：更改动画的初始条件，虚拟环境中不动的和可动的物体的数目、位置等。

### 3.3 基于人工生命的因特网提速

清华大学复杂工程系统实验室是我国人工生命理论研究的基地，他们新近成功地研发了“蚂蚁路由算法在因特网上的应用技术”，建立了一个“蚂蚁路由算法”的仿真网络。仿真的网络节点很多，一如大“网”，当点击源节点后，向很远的目标节点发出一个数据包。模拟开始后，屏幕上许多显示成黑色小方块的“小蚂蚁”背着数据包开始传输，一开始它们显得有点乱，四处乱窜，但几秒钟后，“小蚂蚁”们就在众多的节点中找到了最优路径，并把数据很快有序地传输到目标节点。如果这个最优路径阻断，“小蚂蚁”们又很快找到阻断后的最优路径，完成传输任务。“蚂蚁路由”(Ant-Routing)算法使网络的扩充无限升级，也不会出现环路、重发和阻塞等问题，邮件都能迅速地发送接收，而信息也不会无端丢失。那么，究竟蚂蚁是如何提速因特网呢？

在生物界，蚂蚁是一种头脑简单、视力也很不济的小东西，然而它却有着非凡的辨别复杂路径的能力。每个工作日，单个蚂蚁爬出洞窝时并没有自己明确的目

的或该往哪地方去找食,更不明白整个蚁群的活动范围。但是一旦发现食物源,它就会存储特定的画面,并利用连续的画面来帮助自己在走远后能重新找回来(蚂蚁在路径上留下气味),并寻找最短的路径撤回食物。对蚂蚁找路现象有着深刻研究的清华大学复杂工程系统实验室,把人工生命中“蚂蚁路由原理”具体应用到因特网技术上。不久前,清华大学的研究人员在实验室里开发了一种奇特程序,在这种程序控制下,互联网上的信息会自动精确计算路由的长短,寻找捷径,这也是针对网络规模的无限升级,传统的 Internet 体系结构中出现的“路由问题”而设计的。现在的每一个路由器都记录了其他路由器的信息,而信息量就会随着

网络中子网数目的增加而迅速增加,直到路由器不能承受,于是产生了速度太慢和带宽不够等诸多问题,由此环路、阻塞、重发和邮件丢失就不可避免。如在网络上发一个数据包到美国东海岸某城,邮路先由日本再经夏威夷,再到西海岸,再到美国东海岸。若日本的节点链接断线或堵塞时,每个信息包寻找另外的最优路径,而不会全都堵在一条邮路上。也许“信息蚁”还会立即选择一条从澳大利亚通过的路径。这种随机应变不盲从、不死等的方法有效地提高了网络的利用率和传输速度,然而这一切都有赖于小蚂蚁们对路由长短的精确计算。“蚂蚁路由”原理是人工生命最典型的例子,它的用途还远不止于此(图 38-6-16)。



图 38-6-16 “蚂蚁路由”

### 参 考 文 献

- 1 周济等,机械设计专家系统概论,武汉:华中理工大学出版社,1989
- 2 肖人彬,周济,查建中,智能设计——先进设计技术的核心,机械设计,1997(4)
- 3 钱学森等,关于思维科学,上海:上海人民出版社,1988
- 4 路雨祥,陈鹰,人机一体化系统与技术立论,机械工程学报,1994,30(6),1~9
- 5 周济,查建中,肖人彬,智能设计,北京:高等教育出版社,1997
- 6 李健平,徐林林,廖启,智能设计技术,起重运输机械,2003(5)
- 7 欧阳静安,智能设计体系结构的研究,计算机工程与科学,1999,21(2)
- 8 韩晓建,邓家提,产品概念设计过程的知识表达,制造业自动化,1999,21(5)
- 9 张向军,桂长林,智能设计中的基因模型,机械工程学报,2001,37(2)
- 10 吕大刚,王光远,结构智能优化设计——一个新的研究方向,哈尔滨建筑大学学报,1999,32(4)
- 11 Baras J S, Levine W S, Lin T L, Discrete-time point processes in urban traffic queue estimation, IEEE Transactions on Automatic Control, 1979, AC-24: 12~27
- 12 Pappis C P, Mamdani E H, A fuzzy logic controller for a traffic junction, IEEE Transaction on System Man and Cybernetics, 1987, SMC-7(10):707~717
- 13 洪伟,牟轩沁,王勇等,交叉路网交通灯的协调模糊控制方法,系统仿真学报,2001,13(5):551~553
- 14 Choy M C, Srinivasan D, Hybrid Cooperative Agents with Online Reinforcement Learning for Traffic Control, IEEE 2002,1015~1020



- 15 Bingham E. Reinforcement Learning in neurofuzzy traffic signal control. *European Journal of Operational Research*, 2002, 131:232~241
- 16 Patel M, Ranganathan N. IDUTC: An Intelligent Decision-Making System for Urban Traffic-Control Applications. *IEEE Trans. Veh. Technol*, May 2001, 50(3):816~829
- 17 Kosuke Sekiyama, Jun Nakanishi, Isao Takagawa, Toshimitsu Higashi and Toshio Fukuda. Self-Organizing Control of Urban Traffic Signal Network. *IEEE 2001*, 2481~2486
- 18 Bubak, M Czerwinski P. Traffic simulation using cellular automata and continuous models. *Computer Physics Communications*, 1999, 121-122:395~398
- 19 Schreckenberg M, Neubert L, Wahle J. Simulation of traffic in large road networks. *Future Generation Computer System*, 2001, 17: 649~657
- 20 Neumann J V. The general and logical theory of automata. In: AA Taub. J Von Neumann, *Collected Works*. Urbana: University of Illinois, 1963, 288~298
- 21 Wolfram S. Statical mechanics of cellular automata. *Rev Mod Phys*, 1983, 55(3): 601~643
- 22 Packard N H, Wolfram S. Two Dimension Cellular Automata. *J Statis Phys*, 1985, 38(5/6): 901~946
- 23 Domany E. Exact Results for Two-and Three-Dimension Ising and Potts Models. *Phys Rev Lett*, 1984, 52(11):871~876
- 24 Frish L, Hasslacher B, Pomeau Y. Lattice Gas Automata for the Navier Stokes Equation. *Phys Rev Letts*, 1986, 56:1505~1511
- 25 Madore B, Freedman W. Computer simulation of the Belousov Zhabotinsky reatin. *Science*, 1983, 222:615~621
- 26 Suksy A M. Crystal symmetry. *Phys Bull*, 1976, 27:475~480
- 27 Gerola F, Selden P. Stochastic star formation and spiral structure of galaxies. *Astro physical Journal*, 1978, 223:129~140
- 28 Rosenfeld A. *Picture Languages*. New York: Academic Press, 1979, 1~15
- 29 Mitchell M, Crutchfield J P, Limber P T. Evolving cellular automata to perform computations: mechanisms and impediments. *Physica D*, 1994, 75:361~391
- 30 刘慕仁, 孔令江. 用 12bit 格子气自动机模型模拟静止流和 Couette 流的温度分布. *物理学报*, 1993, 42(6):874~879
- 31 吕晓阳, 孔令江, 陈光旨. 一维几串性细胞自动机演化的临界相变现象分析. *物理学报*, 1990, 39(1):1~9
- 32 Biham, et al. Cellular Automata Model for Traffic Flow. *Phys Rev A*, 1992, 46:6124~6132
- 33 Wolfram S. Cellular Automata as models of complexity. *NATURE*, 1984, 311(4):419~424
- 34 Buijn N G. A combinatorial Problem. *Ned Akad Wetcn Proc*, 1946, 49:758~766
- 35 张本祥, 孙博文著. *社会科学非线性方法论*. 哈尔滨: 哈尔滨出版社, 1997
- 36 Smith A R. Simple computation-universal cellular spaces. *J ACM*, 1971, 18:331~342
- 37 Berlekamp E R, Connay J H, Guy R K. *Winning ways for your mathematical plays*. New York: Academic Press, 1973, 255~300
- 38 应尚军, 魏一鸣, 范英等. 基于元胞自动机的股票市场投资行为模拟[J]. *系统工程学报*, 2001, 16(5):382~388
- 39 刘慕仁, 邓敏艺, 孔令江. 舆论传播的元胞自动机模型(1)[J]. *广西师范大学学报(自然科学版)*, 2002, 20(2):1~4
- 40 魏俊华, 王安麟, 陈武. 城市交通信号自组织控制模型的研究. 投稿论文, 2003, 5
- 41 Jyuhachi ODA, Wang Anlin, et. al. Study on Structural Optimization Technique using Evolutionary Cellular Automata (Local Direct Rule Based on New Transition Function). *Transactions of JSME*, 68-665A, 15~20(2002-1)
- 42 王安麟, 王炬香, 刘传国等. 基于 CA 的结构拓扑设计的研究——平面薄板加强筋拓扑形态创成的仿真. *机械强度*, 2001, 23(2):181~186
- 43 刘传国. 基于 CA 的结构形态设计的初步研究: [毕业论文]. 上海: 上海交通大学, 1999
- 44 苏敏. 基于 CA 的城市汽车流自组织方法初探(控制模型与仿真技术的研究): [毕业论文]. 上海: 上海交通大学, 2002
- 47 吕晓阳, 孔令江, 刘慕仁. 细胞自动机的演化与计算理论. *华南师范大学学报*, 1996(2):43~49

- 48 漆安慎. 细胞自动机与自组织过程. 科学(季刊), 41(8):168~172
- 49 谷永茂, 段国林, 齐红威. 人工神经网络专家系统智能设计平台体系结构的研究. 河北工业大学学报, 2000, 29(4)
- 50 刘浚利, 韩洪成, 李金军. 智能 CAD/CAM 一体化系统研究. 航天工艺, 1999(3)
- 51 钱学森, 于景元, 戴汝为. 一个科学新领域——开放复杂巨系统及其方法论. 自然杂志, 1990, 13(1):3~10
- 52 戴汝为. 从定性到定量的综合集成技术. 模式识别与人工智能, 1991, 4(1):5~10
- 53 戴汝为. 关于智能系统的综合集成. 科学通报, 1993, 38(14):1249~1256
- 54 戴汝为, 王珏, 田捷. 智能系统的综合集成. 杭州: 浙江科学技术出版社, 1995
- 55 王寿云, 于景元, 戴汝为等. 开放的复杂巨系统. 杭州: 浙江科学技术出版社, 1998
- 56 李厦, 戴汝为. 系统科学与复杂性. 自动化学报, 1998, 24(2):200~207
- 57 周登勇, 戴汝为. 人工生命. 模式识别与人工智能, 1998, 11(4):412~419
- 58 Coolsun. 人工生命技术的技术理念和优势. 金融人工生命论坛.
- 59 张永光. “人工生命”的研究进展. 中国科学院院刊, 2000(3)
- 60 涂晓媛, 陈泓娟, 涂序彦. 计算机动画的人工生命方法. 软件世界, 2000(4)
- 61 陈祥. 神奇人工生命提速因特网. 北京青年报, 2000. 4
- 62 Langton C G. : Artificial Life, Addison Wesley, 1989
- 63 Langton C G. et al. Artificial Life II, Addison Wesley, 1990
- 64 Langton C G. et al. Artificial Life III, Addison Wesley, 1992
- 65 Meyer J A, Wilson S W. From animals to animals Proc. 1st Int. Conf. on Simulation of Adaptive Behavior (SAB91), MIT Press 1991
- 66 Ackley D, Littman M. Interactions between learning and evolution, Artificial Life II, also Video Proceedings, Addison Wesley, 1990
- 67 Booker L. Improving search in genetic algorithms, Genetic algorithms and simulated annealing, Davis L. (ed.), 61 ~ 73, Morgan Kaufmann, 1987
- 68 Cariani P. Emergence and Artificial Life, Artificial Life II Langton C G, et al. (eds.), Addison Wesley, 1991
- 69 Casti J L. Paradigms Lost, 1989
- 70 Crow J F. Genetics Notes. Seventh Edition, Dawkins 76 Dawkins, R.; The Selfish Gene, Oxford: Oxford University Press, 1976
- 71 Dawkins R. The Blind Watchmaker, Longman Scientific & Technical, 1987
- 72 Dawkins R. The Evolution of Evolvability Artificial Life, Langton, D G. (ed), Addison Wesley, 1989
- 73 Dewdney A K. Computer recreations, simulated evolution; wherein bugs learn to hunt bacteria, Scientific American, 104~107, May 1989
- 74 Gould S J. An urchin in the Strom, Essays about Books and Ideas, 1987
- 75 Iba H, Akiba S, Higuchi T and Sato T. BUGS: A bug-based search strategy using genetic algorithms, ETL TR92-8, also Proc 2nd Workshop on Parallel Problem Solving from Nature (PPSN92), North-Holland, 1992
- 76 Iba H, Higuchi T, deGaris H and Sato T. A bug based search strategy for problem solving, ETL-TR92-24, also Proc. 13th Int. Joint Conf on Artificial Intelligence (IJCA193), 1993
- 77 Ray T S. An Approach to the Synthesis of Life, Artificial Life II, Langton C G, et al, Addison Wesley, 1991
- 78 Razenberg G. The Book of L, Springer Verlag, 1986
- 79 Sims K, et al. Panspermia, Artificial Life II, also Video Proceedings, Addison Wesley, 1992
- 80 Smith J R. BUGS: A non-technical report, Dept. of Computer Science, Williams College, 1990. 8. 14
- 81 Smith J R. Designing Biomorphs with an Interactive Genetic Algorithms, Proc. 4th Int. Joint Conf. on Genetic Algorithms (ICGA91), 1991
- 82 Wagon S. Mathematical in Action, Freeman and Company, 1991
- 83 Wilson E O. The insect societies, Belknap Press, Harvard, 1971
- 84 Wilson S W. Classifier systems and the animat problem. Machine Learning, 1987, 2(3)

- 
- 85 Wright A H. Genetic algorithms for real parameter optimization Foundations of genetic algorithms, Rawlins J. E. (ed), 205 ~ 218, Morgan Kaufmann, 1991
- 86 中尾智晴. 最適化アルゴリズム. 株式会社昭晃堂出版, 2000
- 87 伊庭齊志. 遺傳的アルゴリズムの基礎. オーム社出版, 1994
- 88 陈国良等著. 遗传算法及其应用. 北京: 人民邮电出版社, 1996
- 89 王小平著. 遗传算法. 西安: 西安交通大学出版社, 2000
- 90 王安麟. 机械工程现代最优化设计方法与应用. 上海: 上海交通大学出版社, 2001
- 91 王安麟主编. 复杂系统的分析与建模. 上海: 上海交通大学出版社, 2004

# 第 39 篇 并行设计

主 编 严隽琪  
编写人 严隽琪  
蒋祖华  
苏 海  
孟祥慧  
刘 洋

# 第 1 章 并行设计概述

## 1 并行工程概念与方法学

### 1.1 并行工程概念

美国防御分析研究院于 1988 年把并行工程定义为：“对产品设计及其相关过程（包括设计过程、制造过程和支持过程）进行并行、一体化设计的一种系统化工作模式。”

并行工程是对产品设计和相关过程（包括制造过程和支持过程）进行集成，开展并行设计的一种系统化方法。换言之，并行工程是在产品设计阶段侧重于同时考虑产品全生命周期（从概念形成到产品回收或报废

处理)中的各种主要性能指标,从而避免在产品研制后期出现不必要的返工与重复性工作。

并行工程曾有许多不同的名称:包括周期工程,生命周期工程,并行产品和过程设计,集成和协同设计,设计合成,生产率工程和系统工程等。无论叫什么,其基本含义都是把产品设计人员和制造工程技术人员结合起来共同改进制造过程和产品。它要求将不同的专业人员(包括设计、工艺、制造、销售、市场、维修等)组成开发小组,亦叫并行开发小组,以协同工作方式进行开放的和交互式的通信联系,以便缩短从概念设计至开始生产的生产准备时间,消除各种不必要的返工,使产品一次开发成功,获得优良的性能价格比。

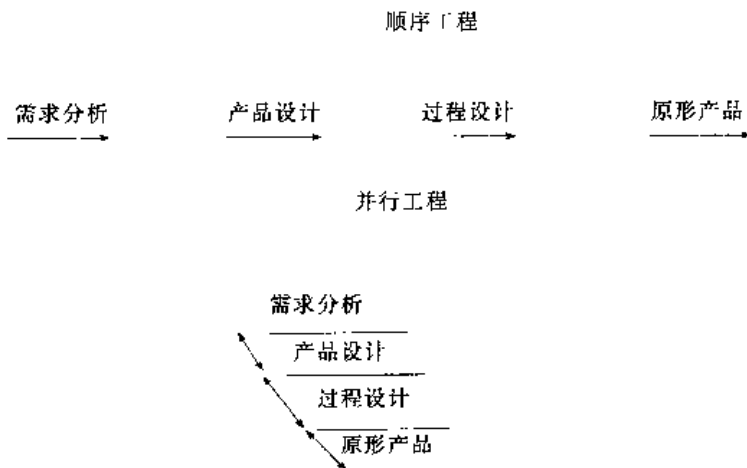


图 39.1-1 顺序工程与并行工程的比较

将并行工程同传统的设计过程作一比较,见图 39.1-1。传统的设计过程可以概括为“顺序工程”或“点到点”的设计。一个设计方案形成并通过后,送给另一个小组(或负责人)修改。每个责任人依次修改设计方案以满足自己的设计目标,产生另一个“点”。这种设计过程的缺点是,每个责任人都代表着特定的专业背景和企业中的某个局部责任,从自己的观点出发,而忽视了其决定对其他责任人的影响。有可能一个责任人作出的修改却不能为另一个责任人所接受,那么就应由初始的责任人重新修改设计方案。如此因相互冲突造成不必要的重复设计,设计效率低下,顺序的生产模式,还使得产品生命周期的各个环节(产品设计、生产、分配和维修)相对独立,其结果是工程决策缺乏远见并导致产品成本提高、质量下降。而且,由于信息在过程中是以顺序方式传递的,因此,要是在下游环节

的工作中发现上游环节的工作需要改善,而这种滞后的信息反馈往往造成大的损失。

并行工程以并行的、交互的、协同的工作方式进行产品及其相关过程的设计。各环节在过程的开始阶段就尽可能地协同工作,信息流是双向的、比较及时地进行反馈。每个制造系统实质上都包括了众多的资源和设计活动,在产品开发过程中如能同时规划这些活动,将是企业实现并行运作的关键,这会使企业节约财富,提高产品质量。

并行工程的特点是:致力于满足用户需求,确信质量乃是改进过程的结果,并认为设计、生产和辅助过程的改进乃是整个企业永恒的责任。

一般说来,并行工程可包括四方面的特性:

1) 并行性:产品和设计过程在同一个时间框架内进行。

2) 约束性: 过程约束性被认为是产品设计的一部分, 从而保证设计出的零件易于制作、搬运和装配, 并便于采用简单成本核算的工艺、工装和物料储运方法。

3) 协调性: 产品和过程密切协调以获得有效成本、质量和交货期的最佳匹配。

4) 一致性: 产品和过程的重大决策要征求全组人员的意见并取得一致。

由于上述“四性”的英文均以 C 开头, 所以也称为并行工程的 4C 性。

采用并行工程这种工作模式, 从产品设计的早期开始, 由于注意到设计的规范性和制造的简单化, 因此“全局成本意识”增强了。与以往的顺序生产方式相比, 并行工程方式也许在产品开发的早期阶段投入了较高的资金, 但整个产品开发和生产成本却会降得更低。不仅如此, 还有其他的优点, 如产品质量和可靠性提高了。图 39.1-2 表示了产品的全局成本意识。图中下方的折线表示超过 80% 的开发成本是处于过程下游的环节——生产阶段和运行支持阶段消耗掉的, 而图中上方的折线却表示处于过程上游的环节——概念设计、初步设计和详细设计对产品最终价值的影响程度达到 90%。如进一步分析, 可以看到, 概念设计阶段的费用仅占总成本的 1% 左右, 却使产品价值的 70% 成定局。因此, 图中用虚曲线表示了修改产品设计的可能性从上游到下游是急剧减少的。因为到了下游阶段, 任何细微的修改都需要付出大的代价。并行工程促使早期的产品概念设计和工程设计尽可能周到地考虑后续过程的要求, 即可制造性、可检验性、可装配性、可维护性、可回收性等等。从而达到产品上市及时、质量

好、成本低、易维护的目的。

图 39.1-3 以定量方式表示并行工程方式加速过程的优越性。图中的数字仅仅是一种假设, 但它表示出在串行工程方式中, 各个环节之间往往留有间歇的时间, 由项目负责人进行调整或修改设计, 他们非固定地用这些间歇时间进行冲突的调解, 交换想法以及向最新版本统一。而在并行工程方式中, 间歇时间得到消除, 生产者依靠计算机系统和产品数据管理软件(在下节将详细论述)保持项目的正常运行。项目管理的许多功能, 包括项目调度、冲突的调解、文档工作、通信、赋权、信息流等均可在计算机上进行。当然这是一张理想化的示意图, 因为工程师 A、B、C、D 的工作不可能是同时开始, 又同时结束的, 但至少他们可以阶梯型开始工作(如图 39.1-1 所示)。

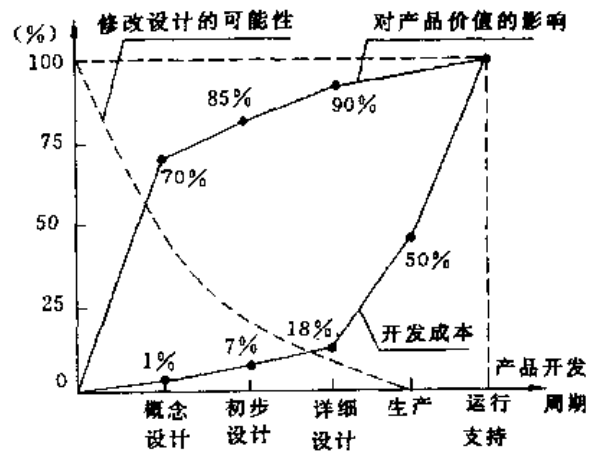


图 39.1-2 全局成本意识

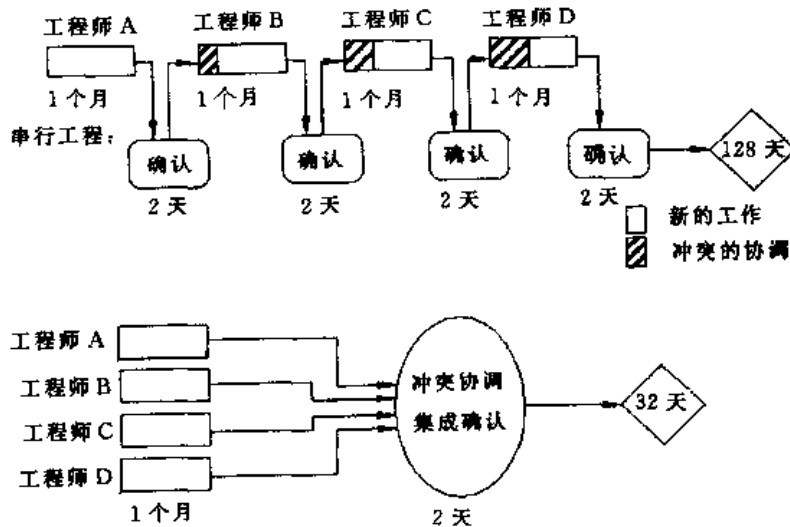


图 39.1-3 并行工程加速设计过程

## 1.2 并行工程方法学

并行工程以信息集成为基础，逐步向产品开发过程集成发展。它继承了以往制造技术中的很多精髓，同时又具有独特的理论和方法学，其要点为：

1) 并行工程的效益来自组织模式改革、用户需求定义、设计环境建设和产品开发新技术应用等四个要素的综合平衡。

根据具体的产品对象以及市场的影响，上述四个要素及其相关关系都在不断地发生变化。只有达到优化意义上的平衡，并行工程才能发挥最大的效益。

2) 从信息集成到产品开发过程集成需要逐级演进。

从信息集成发展到并行过程，经历了五个阶段：互操作的计算环境；互操作的工具和任务；自动化的数据管理；自动化过程管理；自动决策支持。从信息集成到产品开发过程的自动集成体现了重视知识、协作和人的有效参与，并且是从单纯的信息集成向人一管理一技术三者的集成发展。

3) 组织模式的变革经历了三个阶段。

第一阶段是按功能部门划分的传统模式，如市场部、工程部、制造部、销售服务部等，第二阶段是按专业划分小组，如机械组、电子组、动力组等。波音 777 飞机的开发被认为是并行工程方式的成功范例，波音

777 的并行产品开发正是采用了这种模式。第三阶段是按产品机构划分的集成产品开发小组，对飞机而言，如机翼小组、机身小组、推进器小组等。在成功地开发了 777 后，波音公司 1994 年起开始实行集成产品开发小组，亦称多功能小组的方式。

## 2 并行设计特点与本质

### 2.1 并行设计特点

并行设计作为一种设计“哲理”，是在原有信息集成基础上，集成地、并行地设计产品。并行设计更强调功能上和过程上的集成，在优化和重组产品开发过程的同时，实现多学科领域专家群体协同工作。所谓并行，是指一个以上的事件在同一时刻或同一时段内发生，多个复杂性事件可以表示为空间的复杂和时间的复杂，并行设计的目标是尽可能减少时间，通常是采取增加空间的复杂性来实现。从产品设计来看，并行设计采用每个时刻可容纳的设计过程相应增加，使整个设计过程尽可能同时进行。并行设计的特点是并发和协同，并发是指设计活动的并发进行；协同是指多学科设计队伍活动的协作。

国内外学者在研究了大量的设计实例后，提出了图 39.1-4 所示的四种重叠模型，以期其能对产品的并行设计起指导作用。

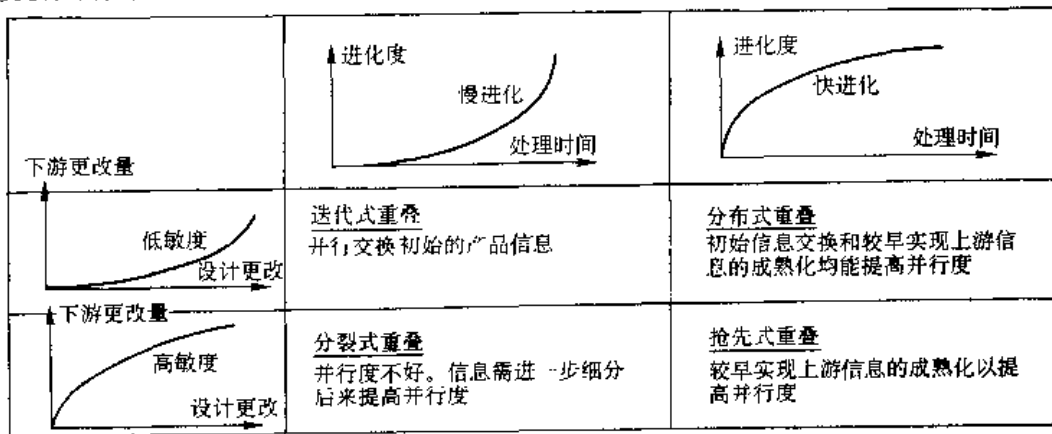


图 39.1-4 基于早期信息交换的四种并行模型

四种重叠模型是基于上游活动参数进化规律和下游活动对参数改变的敏度来建立的。上游参数活动进化规律是指上游活动中涉及的参数从初始值转变为最终值这一迭代过程中的变化规律。下游活动对参数改变的敏度是指当上游活动中的参数变化时，与之相关的下游活动的参数变化率。

这四种重叠模型的比较为：

1) 迭代式重叠模型 (Iterative Overlapping)：下游阶段的参数敏度较低，此时即使是上游的参数变化较

大，对下游的影响也较小，因而可以较早地利用上游的初始信息来触发下游活动的开展。由于此模型中上游活动的参数变化较慢，因而下游活动需要随上游参数的变化不断迭代而逐渐趋于最终值。

2) 抢先式重叠模型 (Preemptive Overlapping)：下游阶段对参数变化敏度较高，而上游的参数进化速度较快，此时交换的信息较早地趋于最终值。这种方式也可以较早地开展下游活动以促使产品开发周期缩短。但由于上游活动结束较早，使其在产品开发的质量控

制方面丧失一些能力。

3) 分布式重叠模型(Distributive overlapping):上游参数进化较快,下游阶段的敏感度较低。此时可以较早地利用上游信息来触发下游活动,而且下游活动也可较早趋于最终值。这是四种模型中最理想的一种并行产品开发方式。

4) 分裂式重叠模型(Divisive overlapping):上游阶段的参数进化较慢,而下游阶段的敏感度较高。此时不适合于用上游的初始信息来触发下游活动,下游活动中的参数也不可能较早达到一稳定值。在这种模型中,需要将所交换的信息进行离散化,依据各个参数的变化规律来确定是否应为早期交换信息。

基于早期数据交换的并行设计方法对产品的设计过程中参数的进化规律及变化敏感度的可预测性要求较高,这两者是实现并行设计的重要基础。基于早期数据交换的并行设计方法更适合于成熟的或有预见性的产品开发过程。

## 2.2 并行设计本质

为了提高产品设计过程的并行度,不能仅仅关注过程信息关联度,还要关注各个阶段产品的信息关联度,以及组织和设计资源的支持程度。

产品数据是在产品设计过程中所需要的或产生的、与产品相关的各种类型的中间数据和最终数据。这些数据可以分为描述产品对象的产品定义数据(如产品的性能数据、几何数据)、产品设计的过程数据(如产品演变的工程数据),以及产品生命周期的产品支持数据(如各种技术标准、规范、标准件和通用件数据等)。它们可能以结构化或非结构化的形式存在于产品开发过程的不同阶段中,并能够被不同的设计活动所共享,这些数据提供对设计活动不同层次上的支持。产品信息视图包括:产品的几何、拓扑表示;产品结构配置、版本等。

产品设计过程的描述不仅仅是多个活动的过程,而且是组织和管理的过程。活动中的组织对象包含了对设计组织的需求,所以设计活动的分解变化,也影响着设计过程中组织的变化。组织是为了完成某种具体目标而建立的社会实体,其结构应同它所完成的设计活动相适应。一个比较完整的组织视角应包括成员与成员之间的关系、角色与角色之间的关系、成员与角色之间的关系、角色与活动之间的关系、角色与资源之间的关系、角色与产品数据之间的关系。按照并行工程的需求,设计组织一般由多个集成化产品开发小组(IPT)组成。组织结构视角包括:组织结构、IPT成员的项目归属、部门归属和群组归属;项目、部门、群组组合人员之间的关系;组织的权限分配。

设计过程中的资源是指保证整个产品设计过程能够顺利进行的硬件设备和软件工具的类型和属性的集合。它包括:数据对象,如设计数据以及技术文档;组织对象,如设计人员等设计过程中需要共享的对象;方法对象,如产品设计中指导设计者设计和决策的各种方法和规则;以及设计所需的硬件设备。对于活动执行顺序的安排比较敏感,在资源规划中,应力求避免或减少冲突的可能性。资源的视图必须反映资源的编号、名称和数量;资源的分布;资源的状态、资源的分类及其调用方法。

产品的并行设计一方面需要增强设计部门、设计小组、设计人员间的信息共享,打破原来串行设计时部门间信息阻滞的壁垒;另一方面需要在设计的初始阶段就全面地考虑下游的可制造性、可装配性、质量保证、环境危害等多种因素,将原来串行设计时的设计修改大循环,改变为不同的设计活动间及同一设计活动中的一系列微小循环。因而并行设计中的难点和重点是合理确定各个设计阶段中设计过程的粒度和设计信息的粒度,使其耦合内容尽早暴露。

从本质上看,并行设计是综合考虑产品信息、产品设计过程信息、组织信息、资源信息等方面,分析设计活动单元实施的条件,将隐性冲突尽早地转化为显性冲突并加以处理,从而减少设计开发中的大返工,力争设计制造一次成功,以缩短产品开发周期,减少开发成本。

## 3 并行设计中的关键技术

并行设计要解决产品设计过程的描述(建模)、优化(重组)、控制(自动化)等方面的问题,而这几个方面都必须满足集成的需要。

### (1) 设计过程的描述及建模

过程模型是一种通过定义组成活动及活动之间逻辑关系来描述工作流程的模型。对流程的描述需要提供逻辑顺序结构,如顺序、分支、汇合、条件、循环、并行的描述。并行设计的描述及建模工作内容中首先要细化分解工作流程,详细描述其数据流动,对每个活动的实施者根据实际需要确定完成该活动所需要的输入信息,并制定该活动的输出信息。对所有活动的信息进行平衡,保证每一输入信息与其对应。同时定义每个成员的具体任务,根据实际情况确定各个活动的开始实施时间和结束时间,详细描述其完成方法及利用的工具。还要根据各信息所涉及的活动及其时间属性,制定出并行设计过程的进程。

### (2) 产品设计过程的优化

过程优化,必须满足产品上市快、质量好、成本低和易回收这一企业和社会期望的基本目标。过程的集



成还不是最终的目的。通过优化以取得产品设计与设计过程满意的性能结果才是目的。优化内容主要涉及产品设计性能与产品可生产性的综合平衡。对过程优化的研究集中于设计方法学的研究、设计过程的质量控制以及过程优化决策的自动化。

在设计方法学的研究中,面向制造的设计(Design For Manufacturing,简称DFM)、面向装配的设计(Design For Assembly,简称DFA)等DFX技术,针对产品生命周期中的不同阶段对候选设计进行评估,是目前并行工程研究的一个十分活跃的领域。过程的“质量功能配置”(Quality Function Deployment,简称QFD)正是从过程的总体上反映了这一概念。

### (3) 过程的控制

过程控制包括了设计过程质量控制及过程优化决策的自动化。所谓质量控制,就是要求制成的产品与最初产品设计定义相吻合,尽量降低由于产品设计过程与后续过程间互相冲突而导致的产品定义修改等等,寻求能使早期产品设计达到尽可能完善的一种实现机

制。这一点需要在早期产品设计过程中,应用前文提及的设计方法学,通过与后续相关过程共享产品模型信息,并作适当的决策才能达到。目前关于设计过程优化的决策,主要还只能依赖于专家的领域知识、经验和科学计算相结合。因此,产品的并行设计,也是人机交互和人工智能相结合的设计系统,由于多知识源之间互相关联、冲突、约束的关系,在决策过程中,协调冲突约束的机制将成为系统运行的管理核心。过程控制应传递决策指令,提供调整系统运行状态的机制,评估系统现行运行状态。对控制的支持依赖于控制结构和网络通信,可包括:任务管理、工具存取和共享管理、多媒体信息交换、远程表示能力等。为此,发展设计过程“质量配置”(QFD)的研究,发展计算机对协同工作支持的系统结构成为实现过程控制的基础任务的两个方面。如目前正处于发展中的“多代理人”(Multiagent)结构,实质上是模仿多个自主决策中心间互相协同工作概念的一种系统结构,正形成当前的研究热点。

## 第 2 章 并行产品开发过程建模及冲突的预消解

### 1 并行产品开发过程模型

#### 1.1 并行产品开发过程的定义

产品开发过程是指某产品从(形成)概念到正式投产之前这一工作流程。

支持并行工程的产品开发过程,简称并行产品开发过程(Concurrent Product Development Process, CPDP),是多功能小组在计算机软硬件工具和网络通信环境的支持下,考虑产品及其全生命周期中有关信息,以缩短产品开发时间,提高产品开发一次成功率为目标而设计的开发活动流程。

在此定义中,跨学科的多功能小组和支持环境是并行产品开发得以顺利进行的基本保证。活动流程是一系列逻辑相关的活动的集合,是并行产品开发过程的核心。活动间的信息流关系、多功能小组的组织关系以及资源约束等是设计该流程必不可少的背景资料。换言之,对并行产品开发过程的完整描述除了开发活动流程外,还应辅以对产品数据、组织关系和资源约束的描述。

#### 1.2 并行设计过程的基本元素

提炼基本元素是实现并行产品开发过程进行形式化描述的基础。现将活动(Activity)、成员(Person)、角色(Role)、资源(Resource)和产品数据(Product

Data)作为其基本元素,其中活动隐含了功能目标与约束。

##### 1.2.1 活动(Activity)

产品开发中的活动有狭义与广义之分。狭义的活动基本上等同于任务,可以描述为“对某对象的处理”,这儿采用的是广义的活动的概念,它在定义时需要同时指定组织、资源和数据三个方面的内容。表达为“某成员(以某角色的名义),用某种资源,完成对某数据对象的处理”。进一步地,可给出如下定义:

活动是指在特定的目标驱动、组织形式和资源保证下完成的对产品数据的处理。

更为一般地,可将活动A抽象为输入流、控制机制、支持机制和输出流组成的四元组,并表示为

$$A = (In, Con, Sup, Out) \quad (39.2-1)$$

其中,In:活动的输入流,表示执行活动所必备的输入数据。一个活动可以有若干个输入流,它们可以来自数据对象或其他任务的输出流。

Out:活动的输出流,表示活动执行后所产生的输出数据。一个活动可以有若干个输出流,他们来自数据对象。作为输出,他们可供应给其他活动。

Con = { Con 1, Con 2, ..., Con m } 表示活动的控制机制(如触发机制),它隐含了功能目标与约束。

Sup = { Sup 1, Sup 2, ..., Sup n } 表示活动的支持机制,它包含组组织对象和资源对象。

上述定义可用面向对象的表达方式描述如下:

```
Class Activity {
    int Activity_Id //活动号
    string Activity_Name //活动名称
    string Activity_Description //活动描述
    int Activity_Status //活动当前所处的状态,分为三种:未执行、执行中和完成
    Function Execute_Controller (Controller 1, Controller2, ...) //活动的控制机制
    Function Execute_Controller_Relationship (Relationship) //活动控制机制间的关系
    Function Execute_Person ( Person_Id1, Person_Id2, ...) //活动的执行人员
    Function Execute_Role ( Role_Id1, Role_Id2, ...) //活动的执行角色
    Function Resource ( Resource_Id1, Resource_Id2, ...) //活动所需的资源
    Function Activity_Data_In ( Data_Id1, Data_Id2, ...) //活动的输入数据
    Function Activity_Data ( Data_Id1, Data_Id2, ...) //活动处理的数据
    Function Activity_Data_Out ( Data_Id1, Data_Id2, ...) //活动的输出数据
    ...
}
```

具有以下特性:

**层次性和分解性:**层次性使得处于高层的活动对其下层活动具有广泛的指导和约束能力;而分解性表明一个复杂的设计活动可以进一步分解为更具操作性的相对简单的活动。

**组织和资源关联性:**对一定的组织和资源而言,只有限定在组织和资源边界内的活动才可能是有效活

动,超出该边界,活动就失去了意义。

### 1.2.2 成员 (Person)

作为活动的执行主体,组织中的人员,即组织的成员,在并行产品开发过程中扮演着极其重要的作用。一般地,它所具有的属性可用面向对象的表达方式描述如下:

```
Class Person {
    int Person_Id           //成员号,一般为工号
    string Person_Name     //成员名称
    string Person_Description //有关该成员的说明资料
    Function Person_In_Team (Team_Id1, Team_Id2, ...) //成员所在的小组
    Function Person_In_Project (Project_Id1, Project_Id2, ...) //成员所参与的项目
    Function Person_In_Department (Department_Id) //成员所属的部门
    Function Person_Is_Role (Role_Id1, Role_Id2, ...) //成员所承担的角色
    ...
}
```

在上述模型中,有“Role (角色)”属性,它相当于一般企业中的岗位。将成员与角色有机地结合起来,构成组织结构的核心部分,是并行产品开发过程组织结构的重要特点之一。

### 1.2.3 角色 (Role)

角色是对组织专业化程度的一种规定,是组织结构的基本单位之一,它必须具备两种属性:

**横向专业化:**对角色所具有的专业知识的规定,是角色执行具体任务的基础。

**纵向专业化:**对角色所具有权利的规定,也就是对通常所说的管理职能的规定。

角色作为组织中专业化程度的规定,是一种抽象的概念,它必须由具体的成员来承担,成员是角色的载体。一个成员可以承担多种角色,一个角色也可以由多个成员来承担。为了更明确地表示角色这一概念,可以给出如下定义:

**概念角色:**当一个角色尚未由具体的成员来承担

```
Class Role {
    int Role_Id           //角色号
    string Role_Name     //角色名称
    string Role_Description //角色说明,如对专业水平的要求,职责范围,仲裁权利级别等等
    string Role_Type     //当前角色类型,分为概念角色和实体角色两类
    Function Role_Person_Can (Person_Id1, Person_Id2, ...) //可胜任该角色的成员
    Function Role_In_Project (Project_Id1, Project_Id2, ...) //角色所属的项目
    Function Role_Person_Is (Person_Id1, Person_Id2, ...) //承担该角色的成员
    (针对实体角色)
    ...
}
```

时,称该角色为概念角色。

**实体角色:**当一个角色由具体的成员来承担时,称该角色为实体角色。

概念角色,在企业中一般表现为岗位空缺。一旦空缺得到填补,即组织中的成员承担了角色,则该角色就转化为实体角色。

概念角色和实体角色的定义将组织设计分为两个阶段:概念角色设计阶段和实体角色设计阶段。概念角色设计阶段根据问题的求解需要设计相应的求解角色,而实体角色设计阶段将概念角色分配给具体的成员来承担,形成实体角色。

角色是体现组织动态性的重要因素,在并行产品开发过程中发挥着不可或缺的作用。角色的介入把原本动静相杂的复杂问题简化为一静(角色关系矩阵)、一动(成员角色关系矩阵)两个问题,使组织能够以相对稳定的基本结构来应对人事的变化,提高了组织的动态适应性。对角色的上述认识可以描述如下:

### 1.2.4 资源(Resource)

资源描述了执行某活动所需的物品、工具或信息等等。广义的资源定义可用图 39.2-1 表示。

由于并行产品开发过程是以人为主体的设计活动,人在并行工程中占有重要的地位,人力资源作为组织构成的基础而在上节以成员与角色的方式予以描述;信息资源既是触发产品开发活动的主要因素,又是

活动结束的主要成果,故将其列入产品数据而在下节详述。因此,在这一节中并行产品开发所需要的资源主要为:

硬件设备:指产品开发所需的各种硬件,如计算机、绘图仪等;

软件工具:指产品开发所需的各种通用工具(如 CAD、CAE、Office 等)和专用支持工具(如油泵仿真设计系统等等)。

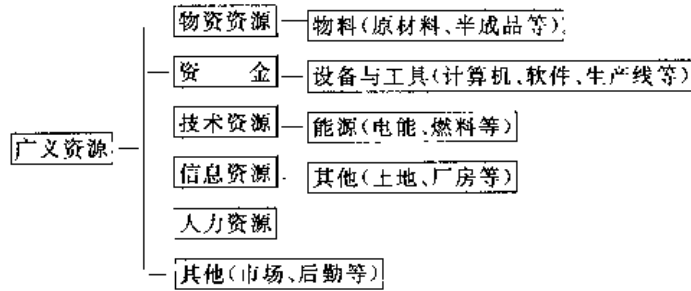


图 39.2-1 广义资源定义

并可给出如下定义:

并行产品开发过程中所涉及到的资源是指保证整个产品开发过程能够顺利进行的硬件设备和软件工具的类型和属性的集合。

设资源空间为  $Re$ , 可描述为

$$Re = (Tr, fr) \quad (39.2-2)$$

其中  $Tr$  —— 资源的论域,对于不同的资源类型,其

论域是不同的;

$fr$  —— 论域的属性,也就是某种资源类型所对应的属性,如完好程度、占用情况等等。

在并行产品开发过程中,更为关心的是资源的竞争性问题。对竞争性资源在难以及时补充的情况下,要结合开发活动作好规划与调度。

上述定义可用面向对象的表达方式描述如下:

```

Class Resource {
int Resource _Id //资源号
string Resource _Name //资源名称
string Resource _Description //资源说明
int Resource _Volume //资源可用数量(对竞争性资源而言)
Function Resource _Type( type1, type2) //资源类型, type1={开发设备,开发工具}
// type2={竞争性资源,非竞争性资源}
...
}

```

### 1.2.5 产品数据(Product Data)

产品数据指在并行产品开发过程中所需要的或产生的与产品相关的各种类型的中间数据和最终数据,这些数据具有以下特点:

数据与活动紧密关联:从开发过程的角度看,产品数据与开发活动之间具有紧密的耦合关系,即一方面,开发活动导致了新的产品数据的产生,或旧产品数据的状态或版本的变化,或视图的变化;另一方面,产品数据又是促使产品开发活动开始和表示产品开发活动

结束的标志。定义并维护二者之间的有机联系,以达到数据与过程的融合在并行产品开发中显得尤为重要;

数量大,类型多:产品开发,特别是复杂的产品开发将产生大量的不同类型的数据。这些数据可以分为描述开发对象的一般产品数据(如产品模型、BOM 表、设计文件、计算书、工艺文件和 NC 程序等)、产品开发的过程数据(如任务单、通知书、会签意见、变更和修改通知单等)和产品支持数据(如各种技术标准、规范、标准件和通用件数据等)。它们以结构化或非结构化的形式存在于产品开发的不同阶段中,并能够被不同的开

发活动所共享。这些数据对开发活动提供不同层次上的支持。

需要良好的组织：上述庞大、繁杂的数据必须以某种良好的组织方式组织起来，即建立不同数据之间的连接关系，才可能给开发活动提供良好的支持。

产品数据的这些特点和由此导致的需求正是产品数据管理(Product Data Management, PDM)产生的背景。目前，主流的 PDM 软件都提供了良好的电子仓库服务功能(Vault Service, 包括数据存储、检索、修改、

浏览及版本管理等)、产品结构管理功能(Structure Management, 如以 BOM 为核心的数据管理)、产品配置管理功能(Configure Management, 提供不同产品结构的不同视图和描述),使得 PDM 成为并行产品开发中不可或缺的支撑平台。

关于产品数据管理更为详细的介绍见第 4 章,此不赘述。

基于以上对产品数据的认识,可将其要点描述为:

```

Class Product_Data {
int Product_Data_Id //产品数据号
string Product_Data_Name //产品数据名称
string Product_Data_Type //产品数据类型
string Product_Data_Description //产品数据说明
string Product_Data_Creator //产品数据的产生者
string Product_Data_Status //产品数据的状态,如产生中,校核、审批、归档等
string Product_Data_In_Vault //产品数据的存储
...
}
    
```

### 1.3 并行设计过程的基本视图

#### 1.3.1 并行设计过程基本元素之间的关系——子视图

本节将元素与元素之间的相互关系表达为各种子视图。这些关系可见表 39-2-1。

由于元素之间的关系是双向的,因而该表具有对称性。本表只标出上三角部分。

某些元素之间的关系可由其他元素之间的关系推导而来,本表中用“+”来表示这种关系。如“活动”与“成员”的关系可由活动与角色的关系(活动分配矩阵)和角色与成员的关系(成员角色矩阵)推出,表示为“活动分配矩阵+成员角色矩阵”。

在活动元素中已明确定义了活动的输入数据、处理数据和输出数据,即产品数据与活动的关系已包含在活动之中,因此资源与产品数据的关系也就可由资源与活动的关系而得到。

显然,只需定义 11 种关系(即子视图)就可明确表示出并行产品开发过程五种基本元素之间的相互关系。这 11 个子视图绝大多数以矩阵的形式表示,其中数据关系矩阵就是产品数据模型。

这 11 个子视图的分类与组合就构成了过程的基本视图集,基本视图集包括活动流视图、组织视图、资源视图和产品数据视图。

由此,得到并行产品开发过程的基本视图集及其所包含的子视图如表 39-2-2 所示。

表 39-2-1 产品开发过程基本元素之间的关系

元素	活动	成员	角色	资源	产品数据
活动	活动网络图	活动分配矩阵 +成员角色矩阵	活动分配矩阵	资源需求矩阵	(包含在活动中)
成员		IPT(集成产品开发团队)结构图	成员角色矩阵	资源使用权限矩阵 +成员角色矩阵	数据访问权限矩阵 +数据存储矩阵 +成员角色矩阵
角色			角色关系矩阵	资源使用权限矩阵	数据访问权限矩阵 +数据存储矩阵
资源				资源关系矩阵	(资源需求矩阵)
产品数据					数据关系矩阵

表 39.2-2 并行产品开发过程的基本视图集

基本视图集	所含子视图
活动流视图	活动网络图
组织视图	IPT 结构图、角色关系矩阵、成员角色矩阵、活动分配矩阵、数据访问权限矩阵、资源使用权限矩阵
资源视图	资源关系矩阵、资源需求矩阵
产品数据视图	数据存储矩阵、数据关系矩阵

1.3.2 活动流视图

活动与活动之间的关系,构成了活动流视图。正如

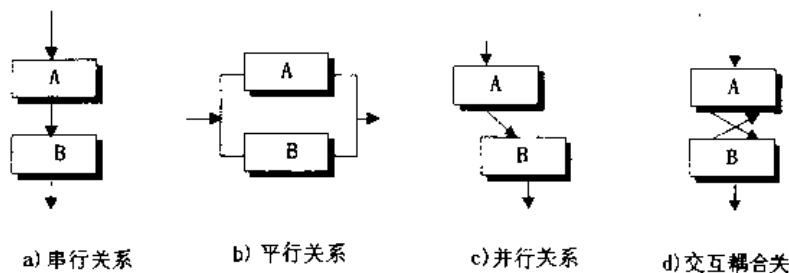


图 39.2-2 活动关系分类

其中

串行关系表示一个活动必须至少得到另一个活动的最后输出信息后才能开始,表现为活动串行开展;

平行关系表示若干个活动之间相互独立,无信息交互,表现为活动可同时开展;

并行关系表示一个活动在得到另一个活动的部分输出信息后就可以开始(该部分信息不是在活动的最后阶段产生),表现为活动异步开展;

交互耦合关系表示活动间存在双向的信息联系,即活动 A 的进行需要活动 B 的信息,同时,活动 B 的进行也需要活动 A 的信息,表现为需经过多次的迭代、反复才能完成。

必须强调的是,活动之间的关系并不是绝对的,而

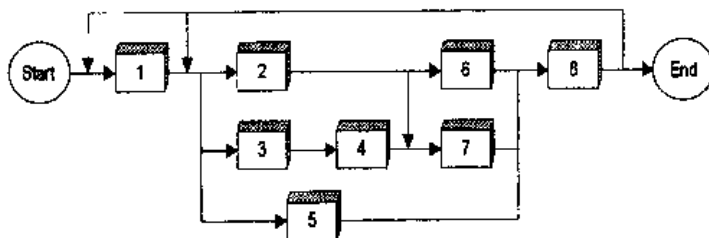


图 39.2-3 喷油泵开发过程的活动流视图

- 活动 1: 柱塞偶件关键参数设计
- 活动 2: 出油阀设计
- 活动 3: 柱塞偶件设计
- 活动 4: 柱塞弹簧设计
- 活动 5: 喷油泵其他零件设计
- 活动 6: 出油阀弹簧设计
- 活动 7: 凸轮设计
- 活动 8: 喷油泵装配仿真

活动是并行产品开发过程基本元素的核心,活动流视图也是并行产品开发过程基本视图的核心。

活动流是描述整个并行产品开发过程全部活动集合上的一个关系。

即设活动集合为 A,则活动流 R 是

$$R = \{(x, y) | x \in A, y \in A\} \quad (39.2-3)$$

其中,  $x \in \phi, y \in \phi$ 。

在并行产品开发过程中,各活动在空间、时间或逻辑上存在着不同的联系。对于不同联系类型的活动在处理上需要采取不同的方式。现从并行产品开发过程建模的需要出发,将活动关系分为以下四类,见图 39.2-2。

会随着活动粒度的划分变化而变化,这就为活动重组从而提高并行度提供了理论基础。

活动间复杂关系的表达一直是一个难题。传统的项目管理一般采用确定型活动网络图,但这种网络图中不允许有任何强连通成分,故只能表达出活动之间的先后时序关系,而不能反映活动间的并行、交互及反馈关系。还可采用随机型网络来表达并行产品开发过程的活动流程,这种网络节点允许自环和回路,并允许出现概率分支(可进行方案优选)。图 39.2-3 为用这种方法表示的喷油泵开发过程的活动流视图。该活动流视图中,如活动 3 与活动 4 是串行关系,而活动 2 与活动 3、活动 4 是平行关系。

1.3.3 组织视图

相对于工业时代的宝塔型组织结构而言,信息时代组织结构的特征是扁平化,以及由此而来的对市场变化的快速反应。在这种背景下,开发组织的组成策略

以及组成形式成为现代产品开发过程中提高产品开发质量,缩短产品开发周期的重要因素。此外,组织构成不能够仅仅被视为一个组织成员的确定过程。作为并行产品开发的一个环节,组织设计还需要与产品结构、产品开发活动等进行有机的协调和综合。这意味着,组

组织的构成还应考虑到产品开发活动、产品数据、开发资源等约束。因此，一个较为完整的组织视图应包括成员与成员之间的关系、角色与角色之间的关系、成员与角色之间的关系、角色与活动之间的关系、角色与资源之间的关系、以及角色与产品数据之间的关系。

(1) 成员与成员之间的关系——集成产品开发小组 (IPT)

为适应日趋激烈的市场竞争，一种被普遍认同的观点是，开发组织的结构将由传统的强调分工和隶属关系的递阶组织结构向动态的、生物结构型的组织方式转变。较为理想的动态组织结构是扁平式组织结构，但就目前的并行工程环境而言，比较适合的是项目——职能型组织，波音公司在开发波音 737-X 时就采用了这种组织结构，取得了较大的成功，他们将这种组织形式称为集成产品开发小组 (IPT; Integrated Product Team)。这种组织结构既保留了传统的静态的职能部门，又能根据项目组成动态的项目工作小组，很适合目

前状况下的产品并行开发模式。图 39.2-4 为 IPT 的组织结构图。

从图中可以看到，左边是由相对固定的职能部门组成的静态组织，它只是形式上的组织，实际操作则按照右边的动态组织。企业领导通过项目开发部直接领导各个项目小组的负责人。各个项目小组的负责人则领导各个项目小组。这种动态的组织结构有以下特点：

- 1) 打破部门界限，减少控制层次，组织结构趋于扁平化；
- 2) 把不同部门的专业人员组织在一起，信息容易沟通，有助于激发员工积极性，发挥专业特长。
- 3) 对某个项目，便于迅速集成，快速响应，可适应复杂多变的市场环境。
- 4) 企业对主管甚至每个员工授权增大，便于灵活决策，提高效率。
- 5) 项目组随产品生命周期的开始和结束而建立和解体，体现了组织的动态性。

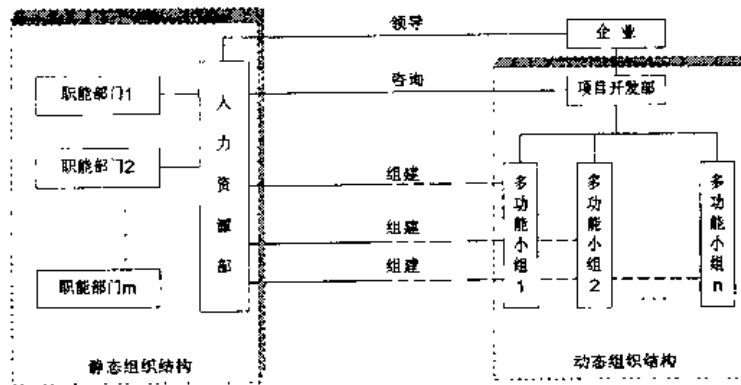


图 39.2-4 IPT 的组织结构图

(2) 角色与角色之间的关系——角色关系矩阵

$M_{Ro-Ro}$

角色与角色之间的关系——角色关系矩阵是组织视图的一个核心内容。角色作为对组织专业化程度的一种规定，必须具有两种属性：横向专业化和纵向专业化。其中，纵向专业化是对角色所具有权利的规定，也就是通常所说的对管理职能的规定，这种规定是一个角色对其上级角色进行服从，对其可控角色进行控制和协调以及对其对等角色进行通讯的基础。角色关系矩阵实际上就是对各个角色的纵向专业化属性进行归纳和抽象的结果，如图 39.2-5 所示。其表达式为

$$(M_{Ro-Ro})_{ij} = \begin{cases} sp & \text{角色 } i \text{ is\_superior\_of 角色 } j \\ jn & \text{角色 } i \text{ is\_junior\_of 角色 } j \\ pt & \text{角色 } i \text{ is\_partnre\_of 角色 } j \\ / & \text{角色 } i \text{ 与角色 } j \text{ 无直接关系} \end{cases} \quad (39.2-4)$$

		Role	
		Role-ID, Name	Role-ID, Name
Role	Role-ID, Name	/	sp
	Role-ID, Name	jn	/
...			

图 39.2-5 角色关系矩阵  $M_{Ro-Ro}$

(3) 成员与角色之间的关系——成员角色矩阵

$M_{P-Ro}$

角色作为组织中专业化程度的规定，是一种抽象的概念，它必须由具体的成员来承担，成员是角色的载体。一个成员可以承担多种角色，一个角色也可以由多

个成员来承担。成员与角色之间的关系可以用成员角色矩阵  $M_{P, Ro}$  来表示,如图 39.2-6 所示,其表达式为

$$(M_{P, Ro})_{ij} = \begin{cases} 1 & \text{第 } i \text{ 个成员可以承担第 } j \text{ 个角色} \\ 0 & \text{第 } i \text{ 个成员不可以承担第 } j \text{ 个角色} \end{cases} \quad (39.2-5)$$

Person	Role	
	Role-ID, Name	Role-ID, Name
Person-ID, Name	0	1
Person-ID, Name	1	1
...		

图 39.2-6 成员角色矩阵  $M_{P, Ro}$

此处的“可以”包含两层含义:

- 1) 该成员有能力承担该角色。
- 2) 小组同意该成员承担该角色,即小组对该成员进行了授权。

角色及成员角色关系矩阵的定义使得原本与人直接相关的问题(如分工、权限等)都转变为与角色相关。我们知道,在并行产品开发过程中,角色是相对稳定的,而成员则由于生病请假或辞职等原因而会有一定的变动。因此,角色的介入把原本动静相杂的复杂问题简化为一静、一动两个问题,使组织能够以相对稳定的基本结构来应对人事的变化,提高了组织的动态性。

(4) 角色与活动之间的关系——活动分配矩阵

$M_{Ro, A}$

活动与角色之间的关系可由活动分配矩阵  $M_{Ro, A}$  (图 39.2-7)来表示。其表达式为

$$(M_{Ro, A})_{ij} = \begin{cases} 1 & \text{第 } i \text{ 个角色承担第 } j \text{ 个活动} \\ 0 & \text{第 } i \text{ 个角色未承担第 } j \text{ 个活动} \end{cases} \quad (39.2-6)$$

Role	Activity	
	Activity-ID, Name	Activity-ID, Name
Role-ID, Name	1	1
Role-ID, Name	1	0
...		

图 39.2-7 活动分配矩阵  $M_{Ro, A}$

(5) 角色与产品数据的关系——数据访问权限矩阵  $M_{Ro, V}$

在并行产品开发过程中,每个小组成员对不同的产品数据拥有不同的访问权限,如查询、修改、复制、删除、归档等等。在组织结构中应定义角色对数据的访问权限,小组成员根据成员角色矩阵中的授权而获得相应的权限。

需要说明的是,产品开发过程中涉及到的数据量很大,如果针对每一种数据定义访问权限,即使是可行的,也非上策。在 PDM 中定义了电子仓库(Vault)这一概念,它是一种逻辑上的划分。因此,本文就通过定义每种角色对不同 Vault 的访问权限,来达到对 Vault 中的数据进行访问和控制的目的。相应的数据访问权限矩阵见图 39.2-8,其中,T(Transfer)表示可在 Vault 间传递数据。其表达式为

$$(M_{Ro, V})_{ij} = \begin{cases} Q & \text{第 } i \text{ 个角色可查询第 } j \text{ 个 Vault} \\ M & \text{第 } i \text{ 个角色可修改第 } j \text{ 个 Vault} \\ D & \text{第 } i \text{ 个角色可删除第 } j \text{ 个 Vault} \\ B & \text{第 } i \text{ 个角色可归档第 } j \text{ 个 Vault} \\ / & \text{第 } i \text{ 个角色不可访问第 } j \text{ 个 Vault} \end{cases} \quad (39.2-7)$$

Role	Vault	
	Vault-ID, Name	Vault-ID, Name
Role-ID, Name	Q	M
Role-ID, Name	Q	/
...		

图 39.2-8 数据访问权限矩阵  $M_{Ro, V}$

(6) 角色与资源的关系——资源使用权限矩阵

$M_{Ro, Rc}$

在并行产品开发过程中,每个小组成员对资源都拥有不同的使用权限,如拥有(Own)、可使用(Use)、不可使用(/)等。在本文的组织结构中,只定义角色对资源的使用权限,小组成员根据成员角色矩阵中的授权而获得相应的权限。资源使用权限矩阵如图 39.2-9 所示。其表达式为

$$(M_{Ro, Rc})_{ij} = \begin{cases} O & \text{第 } i \text{ 个角色拥有第 } j \text{ 个资源} \\ U & \text{第 } i \text{ 个角色可使用第 } j \text{ 个资源} \\ / & \text{第 } i \text{ 个角色不可使用第 } j \text{ 个资源} \\ \vdots & \end{cases} \quad (39.2-8)$$



Role		Resource		
Role-ID, Name	0	U	U	...
Role-ID, Name	/	U		
...				

0: 拥有  
U: 可使用  
/: 不可使用  
...

图 39.2-9 资源使用权限矩阵  $M_{Ro-Rc}$

综上所述,一个完整的组织视图应包括(图 39.2-10):IPT 结构图,成员角色矩阵,活动分配矩阵,数据访问权限矩阵,以及资源使用权限矩阵。

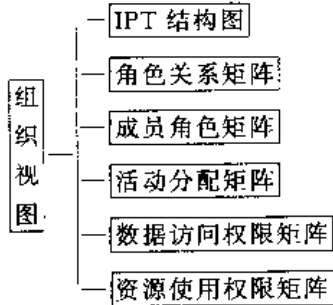


图 39.2-10 组织视图

### 1.3.4 资源视图

并行产品开发过程基本元素中,资源与资源之间的关系及活动与资源之间的关系共同构成了过程的资源视图。

#### (1) 资源与资源之间的关系——资源关系矩阵 $M_{Rc-Rc}$

已知并行产品开发过程中的资源分为硬件设备和软件工具两大类。软件工具必须借助于一定的硬件设备才能起作用,而硬件设备中,如绘图仪与计算机之间也同样存在这种相关性。本文称资源之间的这种关系为资源的依赖关系。进一步地,可将这种依赖关系分为强依赖关系和弱依赖关系。

若某特定资源与另一个特定资源必须同时使用,则称它们之间的关系为强依赖关系,如某型绘图仪与其驱动程序之间的关系。

若某特定资源与另一个特定资源之间虽然相互依赖,但并非必须同时使用,则称它们之间的关系为弱依赖关系,如某绘图仪与某计算机之间的关系。

不同资源之间的这种关系可用资源关系矩阵

$M_{Rc-Rc}$ 来表示,如图 39.2-11。其表达式为

$$(M_{Rc-Rc})_{ij} = \begin{cases} 0 & \text{资源 } i \text{ 与资源 } j \text{ 之间没有依赖关系} \\ 1 & \text{资源 } i \text{ 与资源 } j \text{ 之间是弱依赖关系} \\ 2 & \text{资源 } i \text{ 与资源 } j \text{ 之间是强依赖关系} \end{cases} \quad (39.2-9)$$

Resource		Resource		
Resource-ID, Name	0	2		...
Resource-ID, Name	2	0		
...				

0: 没有关系  
1: 弱依赖关系  
2: 强依赖关系

图 39.2-11 资源关系矩阵  $M_{Rc-Rc}$

#### (2) 活动与资源之间的关系——资源需求矩阵

$M_{A-Rc}$

活动必须由小组成员借助一定的资源才能完成,即活动对特定资源存在需求关系。这种关系可用资源需求矩阵  $M_{A-Rc}$ 来表示,见图 39.2-12。其表达式为

$$(M_{A-Rc})_{ij} = \begin{cases} 1 & \text{第 } i \text{ 个活动需要第 } j \text{ 个资源} \\ 0 & \text{第 } i \text{ 个活动不需要第 } j \text{ 个资源} \end{cases} \quad (39.2-10)$$

Activity		Resource		
Activity-ID, Name	1	0		...
Activity-ID, Name	1	1		
...				

0: 不需要该资源  
1: 需要该资源

图 39.2-12 资源需求矩阵  $M_{A-Rc}$

由于资源与资源之间的关系在资源关系矩阵中已得到定义,因此,在资源需求矩阵中只需注明活动的关键资源,特别是竞争性资源,其他一般性资源可由资源关系矩阵推出。

### 1.3.5 产品数据视图

产品数据视图负责对产品数据的组织和管理,为了达到这一目的,它需要回答两个基本问题:产品数据

放在何处,产品数据与产品数据之间具有怎样的关系。

#### (1) 数据存储矩阵

定义产品数据的存储位置(物理的或逻辑的)是对产品数据进行管理的基础。在创建一个产品数据时,应对其处于不同状态时(如产生中,校核,归档等)的存储位置做出规划,因而数据的存储矩阵是动态变化的。

#### (2) 数据关系矩阵

产品数据与产品数据之间的关系因数据类型不同而有多种形式。一般地,产品数据可分为两大类,产品结构数据(如表示产品结构信息的各种图形数据)和

产品描述数据(如设计文件等字符级数据)。它们之间的关系可用数据关系矩阵描述。由于产品数据视图是产品数据管理的主要研究内容,本节不对其作更深入的展开。

### 1.4 并行设计过程的递阶集成多视图模型

为了全面表达并行产品开发过程,特在上述这些视图的基础上,构造一个递阶集成多视图模型(The Hierarchical Integration Multi\_views Model,简称 HIMM)(图 39.2-13)。该模型具有以下特征:

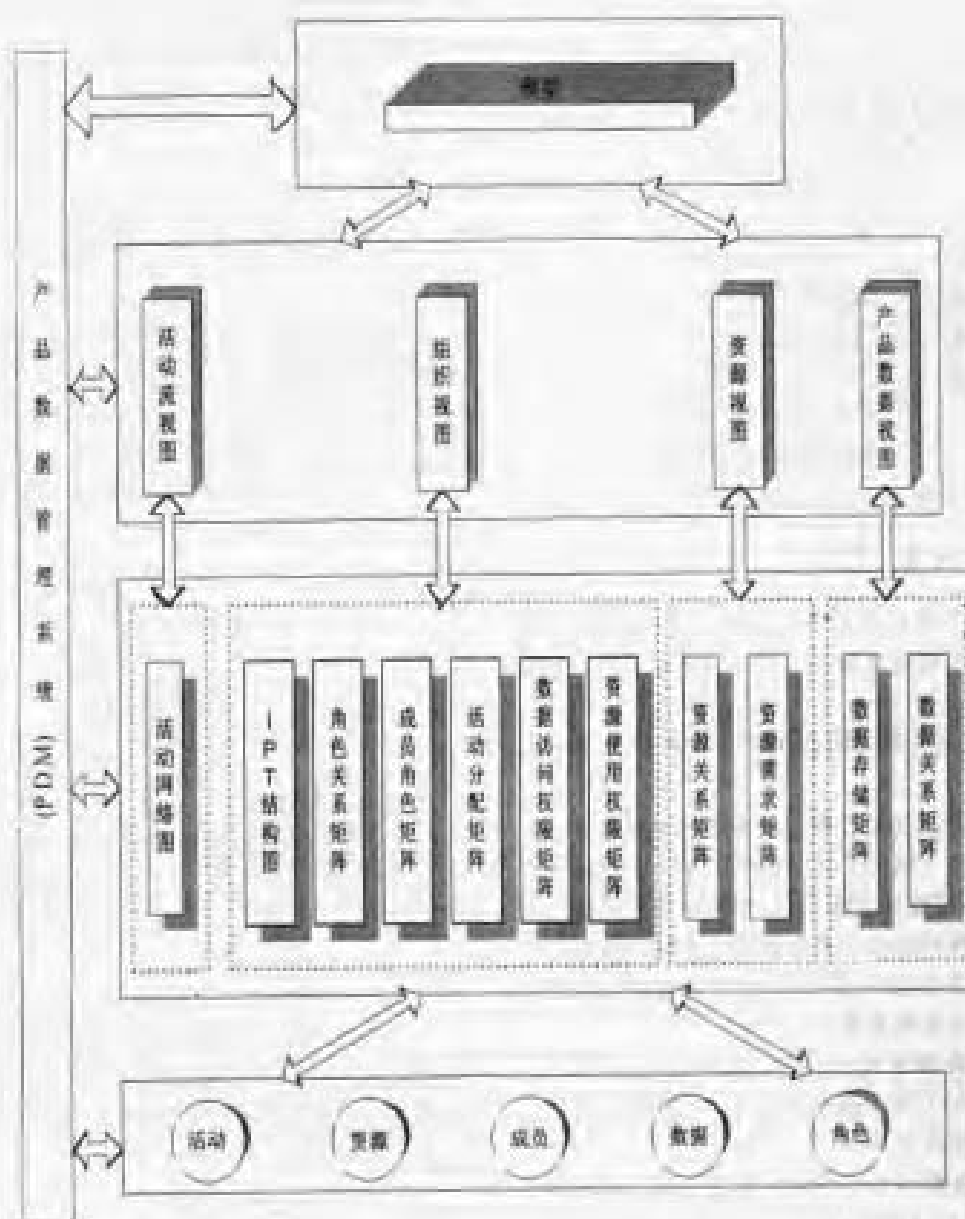


图 39.2-13 并行产品开发过程的递阶集成多视图模型

1) 多视图:视图代表了并行产品开发过程的不同侧面,因而多视图便于实现对过程的全面理解,为过程管理奠定基础。如果缺乏某一视图,则对并行产品开发过程的分析、监控和协调就会因信息不全而难以进行。各视图之间的联系见图 39.2-14;

2) 集成性:并行产品开发过程模型不仅包含了多个视图信息,而且各视图之间互相关联,因此,集成性

特征能够自动维护视图的一致性。产品数据管理系统(PDM)是模型的集成平台;

3) 递阶性:并行产品开发过程的递阶集成多视图模型将模型本身分为四个层次:元素层——子视图层——视图层——模型层。元素与元素之间的相互关系构成了子视图,子视图的集成构成视图,而视图的集成才最终构成模型。

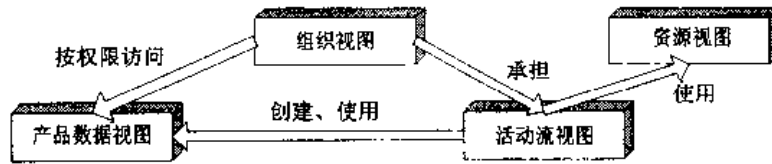


图 39.2-14 并行产品开发过程递阶集成多视图模型的视图联系

## 2 并行设计过程中冲突的预消解

### 2.1 冲突分析

#### 2.1.1 冲突的形成

设计问题实质上可将每个设计活动看作是一个对多个设计变量的约束方程组的求解问题。由于活动是互相关联的,所以有一部分变量会重合,称为这些活动的公共变量。显然其余的变量可以根据他们与公共变量的关系而分为两部分:与一个或若干个公共变量直接相关的变量,称为耦合变量;与任一公共变量都不直接相关的,称为独立变量。

因此,活动的求解可以分为三部分,即对独立变量的求解,对公共变量的求解及对耦合变量的求解。由于公共变量的存在,并行产品开发往往表现为若干个活动的约束方程组的联合求解。如图 39.2-15 所示。

图中  $I_A, R_A$  为活动 A 的独立变量与耦合变量,  $I_B, R_B$  为活动 B 的独立变量与耦合变量,  $C_1, C_2, \dots, C_m$  为活动 A 与活动 B 的公共变量。当两个活动的约束方程组联合求解失败,则意味着两个设计活动之间存在着冲突。

除了设计问题以外,并行产品开发中还可能遇到人员角色定义与活动流程不符(组织冲突),活动所需资源无法得到(资源冲突)等其他冲突,这里就不一一细述。

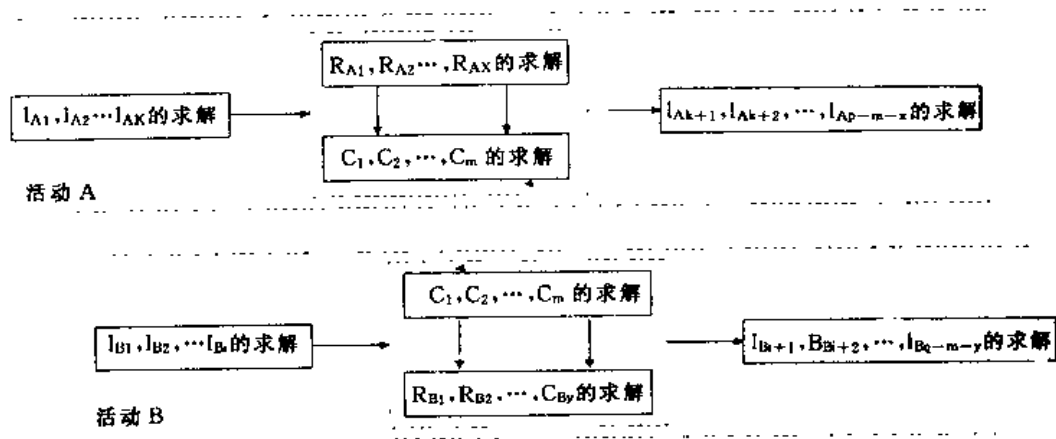


图 39.2-15 并行设计图示

#### 2.1.2 冲突的概念与内涵

Rahim 认为并行工程中的冲突是一种“相互作用状态(interactive state)”,这种状态表现为实体内或实体间的不一致性,或不相容性。这里的实体指开发目标、开发小组、活动或 Agents。由于并行产品开发过程

各个视图之间都可能发生冲突,因而冲突的种类较多。在 Rahim 的定义基础上,可将冲突进一步概括为:

冲突是期望与环境间不和谐关系的一种描述。

该定义中,“期望”与“环境”随冲突类型的不同而有不同的含义。如一个设计者的环境可包括两部分:小

组中其他人的设计选择构成其中的一部分,属于该个人的设计职权范围的各种可选设计参数(结构)构成环境的另一部分。因此设计者的活动可以大致描述为根据设计目标(期望)和其他成员已作出的设计选择,选择个人所能支配的设计参数,或判断自己的设计目标是否已不可避免地受到现有环境的限制而难以实现。

冲突按其性质可分为两种,竞争性冲突和协作性冲突;前者是冲突各方不肯让步的冲突;后者是冲突各方可以让步的冲突。

冲突的竞争性与合作性是相对的,当更高层次的期望值更改后,竞争性冲突就转化为协作性冲突。竞争性冲突既有人为的原因,也有技术的原因,因为技术上也有不可让步的情况存在,此时只能通过仲裁手段来强制冲突中的一方牺牲自己的指标,满足另一方的要求;或通过协商修改更高层次的指标。而协调协作性冲突的策略则更多地与技术有关,如弱化乃至删除对全局不重要的指标,尽可能多地沟通等等。

当然,与人相关的冲突中一个重要的因素是人的社会因素和心理因素,如信任度,成员的自尊心,对合作者的接受程度,从合作者那里感觉到自己的被接受程度等等,这些因素常常会对协作产生一定程度的影响。

并行工程是以集成的并行的方式设计产品及其相关过程(包括制造过程和支持过程)的系统方法,这种方法力图使产品开发人员从一开始就考虑到产品全生命周期中的所有阶段和因素,如用户需求、质量、成本、制造、维护、报废等等,从而保证产品开发的一次成功(right the first time)。为了达到这一目的,并行工程强调通过由领域专家组成的多功能小组的协作来完成产品开发。在协作过程中,由于多功能小组内和小组间各个领域专家在各自的目标、学科背景和领域知识水平上存在差异,导致他们考虑问题的角度不同,评价标准也不一样,因此,在进行设计时作出的选择也肯定会有差异,这就难免产生冲突。同时,由于多功能小组内和小组间存在大量相互依赖的关系,它们的活动也交织在一起,相互影响,相互制约,因活动缺乏良好的规划而引起的冲突也时有发生。冲突可能发生于产品开发的任何阶段。所以,从某种意义上说,产品开发过程既是一个开发过程本身不断向前推进的过程,更是一个冲突不断产生和协调的过程。

冲突对产品开发过程有着一定的负面影响和阻碍作用,但同时也有其积极的一面。冲突是一个革新设计的开始。冲突的产生促使设计人员及早发现问题并解决问题,最终实现提高产品各项性能指标,最大限度地满足各领域专家要求的目标。随着设计中的冲突不

断地被协调成功,一个完善的更优的设计也随之产生。

并行产品开发过程运作的核心是冲突协调。

### 2.1.3 冲突产生的原因

根据冲突形成的机理以及上文对冲突研究范围的界定,可以将并行产品开发中冲突的主要产生原因归纳为以下几种:

1) 不一致的知识表达方法 由于缺乏共同的知识表达方法,缺乏一个共同一致的开发对象模型,一个领域的知识和信息难以被正确地传达给另一个领域的专家。不同领域的专家之间难以进行有效的交流,常因领域知识的误解而产生冲突。

2) 不一致的领域假设 在并行产品开发过程中,设计者常常需要借助于其他各方的开发信息。由于其他方开发工作尚未完成,开发信息不全,设计人员就需要根据以往的经验对其他各方的开发工作作出一定的假设,以完成自己的开发工作。若这种假设出现偏差甚至错误,则当这些局部开发方案集成为一个总体方案时,就会暴露出相互矛盾、互不协调的问题。

3) 不一致的领域目标 活动分解是一个复杂的工作,因为一个活动对应着一个任务,而每个任务都应有自己明确的目标。当总目标分解为子目标时,子目标之间,子目标与总目标之间难免有不明确的乃至错误的关联。而活动的设计者在求解自己的局部问题时,总是尽可能使自己的目标值取得最优,而较少,也较难考虑到其他领域的目标,冲突因此而产生。

4) 不一致的评价准则 由于各领域专家考虑问题的角度不同,对开发结果进行评价的标准不同,在开发过程中常常会作出不同选择,冲突因此而产生。

5) 过程规划的缺陷 如组织视图与活动流视图不一致;集成产品开发团队(IPT)的设计、角色体系设计与活动分解、活动分工不一致,或者与活动流不一致,从而造成活动执行时人员/角色安排上的冲突;或资源视图与活动流视图不一致;活动执行时由于所需资源(不含人力资源)的缺乏而引起冲突。

6) 变化的开发环境引起的“失衡” 并行产品开发过程模型的建立是并行工程的实现的基础。随着产品开发的逐步深入,产品开发的各种资源、技术条件等也都处在变化之中,当产品开发过程模型已无法如实反映实际的产品开发过程时,冲突的产生就不可避免。

上述六种原因中,前四种主要造成数据冲突,而组织视图、资源视图与活动流视图之间的不一致,既可造成组织冲突/资源冲突,又可造成活动流冲突,变化的开发环境使得产品开发过程的各个视图之间的协调性遭到破坏,可能导致各种冲突的爆发。

### 2.1.4 冲突的分类

#### (1) 从冲突发生的范围区分

从冲突发生的范围,可将冲突分为:多功能小组中单个成员内部的冲突,多功能小组中成员之间的冲突和多功能小组间的冲突。

1) 多功能小组中单个成员内部的冲突 即使是一个人完成的开发活动,也会遇到多种权衡、取舍与决策,这种领域专家在完成本领域的开发工作时所遇到的只与本领域有关的冲突,称为多功能小组内单个成员内部的冲突。

2) 多功能小组中成员之间的冲突 多功能小组

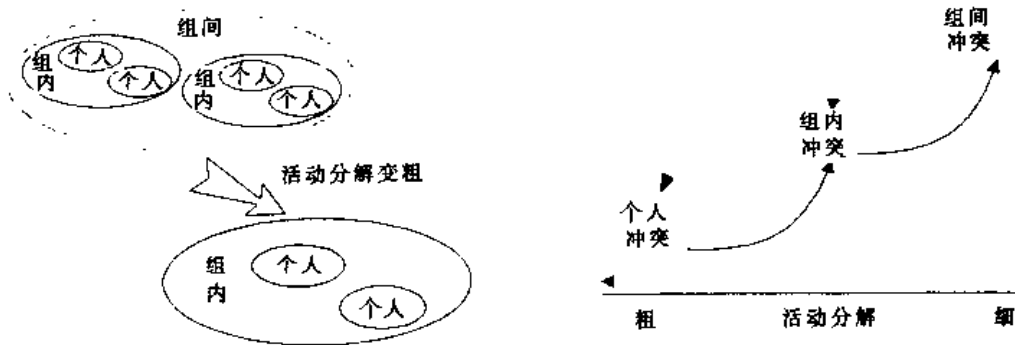


图 39.2-16 冲突类型的转化

#### (2) 从冲突的显示度区分

根据冲突暴露的明显程度,可将冲突分为显式冲突和隐式冲突两类。前者指明显暴露出来的,使活动不能进行下去的冲突;后者指已经存在于开发过程内部,但尚未明显暴露出来,因而活动仍能进行的冲突。

显式冲突与隐式冲突也并非固定不变,一旦隐式冲突随着开发过程的推进而明显暴露,导致开发活动难以进行下去则就转化为显式冲突。从并行工程的本质看,应该是使隐式冲突尽早地转化为显式冲突,以便提早消解冲突。

#### (3) 从冲突的内容区分

并行产品开发过程不同视图对应的冲突具有不同的内容,可以将冲突分为数据冲突,组织冲突,资源冲突和过程冲突四类。

1) 数据冲突 产生于产品数据视图的冲突,如某参数的取值难以达成共识。根据数据冲突所涉及的对象的不同,可将数据冲突进一步细分为:参数冲突,结构冲突,方案冲突三大类。

2) 组织冲突 集成产品开发团队(IPT)的设计、角色体系设计与活动分解、活动分工不一致,从而造成活动执行时人员/角色安排上的冲突。组织冲突产生于并行产品开发过程的组织视图。

的领域专家在完成本领域的开发工作时所遇到的只与本小组其他领域有关的冲突,称为多功能小组内成员之间的冲突,简称为组内冲突。

3) 多功能小组间的冲突 多功能小组在完成小组工作时所遇到的涉及到其它小组的冲突,称为多功能小组间的冲突,简称为组间冲突。

上述三种冲突的类型并非固定不变,它们随着活动分解、角色设计和活动分配工作的粒度不同可相互转化,活动分解的越细,角色体系越庞大,组间冲突相对就越多,个人内部冲突相对就越少,其相互关系可表示如图 39.2-16。

3) 资源冲突 活动执行时由于所需资源(不含人力资源)的缺乏而产生的冲突;造成资源缺乏的原因,一方面是资源本身的匮乏,另一方面则源于对资源缺乏良好的规划与调度。资源冲突产生于并行产品开发过程的资源视图。

4) 活动流冲突 并行产品开发过程的活动流视图与实际开发环境之间的冲突。造成这种冲突的原因在于,产品开发的各资源、技术条件等随着产品开发的逐步深入也会不断变化。活动流冲突产生于并行产品开发过程的活动流视图。

上述冲突之间存在一定的关联,特别是组织冲突、资源冲突和活动流冲突之间。人员/角色安排上的冲突和资源冲突也可能同时使原有活动流变得难以执行,故表面上的多个冲突实际上可能是等价的,或具有因果关系。因而在协调这类冲突时应从分析冲突之间的关系入手,从而选择回报率最高的方法。

## 2.2 冲突的预消解

所谓冲突的预消解,就是通过应用一定的管理手段和技术手段,尽可能早地将隐式冲突转化为显式冲突,从而在一定程度上减少和避免后期出现大的显式冲突造成大的损失。因此,冲突的预消解体现了并行工

程思想。

对冲突进行预消解的原则性方法包括合理的组织设计与活动分解、集中的产品数据管理、领域知识共享、辅助工具的应用和采用面向产品生命周期各/某环节的设计(DFX)等。

#### (1) 合理的组织设计与活动分解

合理的组织设计与活动分解能有效地早期发现并解决冲突,提高并行开发的效率。从组织结构的角度来说,多功能小组是并行产品开发过程的一个基本组织单元,在进行多功能小组的设计和相应的角色体系的设计时,要充分考虑到产品功能结构的分解和这些功能结构对专业水平的限制。合理的活动分解可充分降低各活动之间的耦合度,提高设计人员工作的独立性。而活动的目标不清或设计者的职责不明,是并行产品开发过程中冲突发生的重要原因。

合理的组织设计与活动分解还要保证组织与活动之间的一致性,故在设计完并行产品开发过程模型的组织视图和活动流视图之后,还要从对方的角度对本视图进行考察,尽可能减少过程执行中冲突发生的可能性。

#### (2) 集成的产品数据管理

在并行工程环境下存在着数量庞大的产品数据及相关数据,且大量的数据是在产品的动态开发过程中形成的。各种信息及状态的变化会影响其相关数据的有效性。在缺乏有效的管理手段和技术手段的情况下,工程技术人员难以及时、全面地得到开发过程中的各种数据的变化信息,难以保证当前数据的有效性,冲突的发生也就在所难免。为了避免引起各种冲突,需要产品数据管理框架对并行工程环境下所有的产品信息、过程信息和支持信息进行综合管理和控制,为并行工程提供一个并行化的产品开发支持环境。该环境能有效降低开发过程中的混乱与无序,从而消解冲突的发生源。

#### (3) 领域知识共享

在并行产品开发过程中,一个重要的问题就是不同领域专家对于开发对象的领域知识的共享问题。不同的领域专家对同一个对象的不同侧面进行工作,由于各自考虑的角度不同,专业背景不同,对同一个对象的描述往往存在着很大的差别。同时,由于使用的专业语言不同,不同的领域专家相互之间难以进行交流,一个领域的意见往往无法正确地反映、传达给另一个领域的人员。在并行产品开发过程中,对象的领域知识共享是成功协作的重要基础。因此,在并行工程环境下需要提供一个领域知识共享的环境,在此环境中,各领域的专家能够精确地表达各自的见解,同时正确地理解其他领域专家的知识 and 想法,避免因领域知识的误解

而产生许多不必要的冲突。

#### (1) 辅助工具的应用

预消解的关键是能提前检测到冲突,这需要多功能小组领域专家知识的支持,需要丰富的相关信息资源的储备,需要仿真工具的支持,如用过程仿真工具对规划的活动流进行仿真,从而发现潜在的活动流冲突或资源冲突,同时还需要有效的检测工具的支持,这种检测工具应具备检测隐式冲突的能力。

#### (5) 面向产品生命周期各/某环节的设计(DFX)

DFX 是 Design for X(面向产品生命周期各/某环节的设计)的缩写。其中,X 可以代表产品生命周期或其中某一环节,如装配、加工、使用、维修、回收、报废等,也可以代表产品竞争或决定产品竞争力的因素,如质量、成本、时间等等。而这里的设计不仅仅指产品的设计,也指产品开发过程和系统的设计。

DFX 是一种哲理、方法、手段、工具。首先,所有的产品开发人员都应该认识到,在设计阶段尽早地考虑产品全生命周期将有益于产品竞争力的提高。其次,产品设计人员、产品开发管理人员应该积极应用 DFX 方法于产品设计。最后,借助计算机实现的 DFX 工具可以有效地辅助产品设计人员按照 DFX 方法进行产品设计。

DFX 本身不直接产生设计方案,而是设计评价分析方法,为设计提供依据,DFX 方法的应用最终通过再设计实现产品的优化。DFX 方法不仅用于改进产品本身,而且用于改进产品相关过程(比如,装配过程、加工过程等)和系统。以往当产品设计完毕,紧接着进行过程设计(比如,加工工艺设计,装配工艺设计等),而 DFX 方法强调产品设计和过程设计的并行进行。

#### 典型的 DFX 方法

面向装配的设计 DFA(Design for Assembly)。装配指将零件结合成为完整的生产过程。所谓面向装配的设计是旨在提高装配的方便性以减少装配时间、成本的设计。而向装配的设计原则包括:①减少零件数;②采用标准紧固件和其他标准零件;③零件的方位保持不变;④采用模块化的部件;⑤设计可直接插入的零件;⑥尽量减少调整的需要,等等。

面向制造的设计 DFM(Design for Manufacture)。这里的制造主要指构成产品的单个零件的切削、铸造、锻造、焊接、冲压等冷热变形加工过程。DFM 用于为了减少该类加工的时间与成本,提高加工质量的零件设计评价。面向制造的设计原则主要有:①简化零件的形状;②尽量避免切削加工因为切削加工成本高;③选用便于加工的材料;④尽量设置较大的公差;⑤采用标准件与外购件;⑥减少不必要的精度要求,等等。由于不同的加工方法存在较大的差异,具体到面向不同加工

的设计,必然有不同的具体的设计原则。

面向检验的设计 DFI (Design for Inspection) 着重考虑产品、过程、人的因素以便提高产品检验的方便性。产品检验是加工和维修的主要工作。加工中的产品检验是为了提供快速精确的加工过程反馈,而维修中的产品检验则是为了快速而准确地确定产品结构或功能的缺陷,及时维修以保证产品使用的安全。产品检验方便性取决于色彩(比如电路板上元器件的颜色对应不同种类)、零件内部可视性(比如液压缸等液体容器应该直接显示液面)、结构等等诸多因素。

面向维修的设计 DFS (Design for Service/Maintain/Repair)。产品的售后服务主要是指产品维修,而产品维修主要涉及产品拆卸和重装等工作,因此产品维修性主要取决于产品故障确定的容易程度、产品的可拆卸性和可重装性,也取决于产品的可靠性。维修总是伴随着拆卸和重装,如何减少这些工作所需时间和成本是 DFS 的重要问题。另一方面,考虑零部件的可靠性的结果是要尽量使容易发生故障的零部件处于容易拆卸的位置,从而有利于维修时间与成本的减少。改进产品可维修性的原则包括:①提高产品可靠性;②经常需要维护的零件和易磨损易失效的零件置于易发现和易接近处;③零件的拆卸尽量不需要移动其它零件;④尽量减少需要使用的维修工具种类;⑤模块化设计;⑥采用标准件;⑦留有足够的维修空间,等等。

面向回收的设计 DFR (Design for Recycling) 日渐减少的自然资源、有限的垃圾填埋空间、有害废弃物的危害等现实已经迫使一些工业先进国家制订相关法规条例,促使产品回收开始成为企业的责任。在这种环境下,企业必须将产品回收问题提到日程上来。面向产品回收的设计重点集中在产品的可拆卸性的提高和材料方面。产品的可拆卸性取决于零件数、产品结构,拆卸动作种类、拆卸工具种类等因素。

面向质量的设计 DFQ (Design for Q) 质量是产品的生命。质量可以理解成产品满足要求的程度。提高产品质量贯穿产品生命周期各个环节。面向质量设计的主要的原则包括:①产品应易于检查;②采用标准件;③图纸标注清楚规范;④尺寸公差设置合理;⑤模块化

设计,等等。

面向可靠性的设计 DFR (Design for Reliability)。可靠性指产品在一定的使用条件下在特定时间内令用户满意地实现其功能的概率。可靠性与质量密切相关。不考虑质量的可靠性是无意义的,而只考虑质量而不考虑可靠性的产品缺乏市场竞争力。面向可靠性的设计目的在于创造出具有可靠性的高质量的产品。产品可靠性的指标包括平均失效时间、单位时间内的失效次数等。决定产品可靠性的因素包括产品复杂程度、零件的可靠性、冗余件与备用件的使用、可维修性等等。因此,改进产品可靠性的原则包括:①简化产品结构;②增加排除环境因素干扰的设计(如以密封件避免湿气,以屏蔽罩避免电磁与静电辐射等);③采用标准件和材料;④减少导致疲劳失效的设计,如减少应力集中;⑤紧固件争取采用可锁定的;⑥提高零件的冗余度,等等。

面向环境的设计 DFE (Design for Environment)。DFE 着重考虑产品开发全过程中的环境因素,目的在于尽量减少在生产、运输、消耗、维护与修理、回收、报废等产品生命周期的各个阶段,产品对环境产生的不良影响,如资源衰竭(生物与非生物)、污染(臭氧层破坏、全球暖化、酸雨、噪声等)、失调(干旱、地表变质等)等等。在充分意识到环境因素下开发出来的产品往往不仅对环境产生的不良影响少,而且消耗少、成本低、易为社会接受。因此重视面向环境的产品开发的企业能够具有较大的竞争优势。产品开发中对环境产生较大影响的主要因素包括材料、加工处理、功能、形状、尺寸、配合与安装等。减少对环境不良影响的设计原则包括:①设计可重复使用、可回收的产品及其零件;②减少零件数;③减少使用材料的种类;④尽量少用玻璃、金属强化塑料等复合材料;⑤使用压模铭牌而不使用铭牌片,等等。带有不分离拉环的饮料罐是面向环境设计的代表产品。与面向环境设计同义的术语包括:可持续发展(sustainable development)、生态设计(ecodesign)、绿色设计(green design)、生命周期设计(lifecycle design)、环境意识设计(environmentally conscious design)等。

## 第3章 workflow 管理系统

### 1 workflow 的基本概念

workflow 技术是 20 世纪 80 年代中期逐渐发展起来的。进入 90 年代后，随着企业之间竞争日趋激烈和计算机网络及通信技术的发展，越来越多的企业利用信息技术去支持它们的业务过程处理，以适应快速变化的商业环境以提高企业的工作效率和工作质量。workflow 技术得到了企业界和学术界的广泛重视，成为研究热点。1993 年，由国际著名公司、研究机构和高等院校成立了 workflow 管理联盟（Workflow Management Coalition, WfMC），该联盟的成立标志着 workflow 技术开始进入相对成熟的阶段。WfMC 目前有 280 多个成员组织，遍布世界各地。该组织由三个委员会组成，分别是技术委员会、对外关系委员会和筹划指导委员会。其目的是通过制定 workflow 技术及其标准，提高不同 workflow 产品之间的连通性和协同工作能力。通过使用标准可以使不同的产品之间协同工作，也可以改善 workflow 产品与其他 IT 服务（电子邮件、文档管理）之间的集成。经过该组织的不懈努力，workflow 标准的制定和推广工作进展得非常迅速，目前，多数 workflow 产品的生产厂商已经在产品中遵循了全部或部分标准。它将 workflow 活动分解成定义良好的任务，由具有不同角色的人员按照一定逻辑过程进行执行。

并行工程要求对产品开发过程加强控制和管理，从而使整个产品开发过程能够有序、高效地完成。workflow 技术可以很好地用于产品开发过程的建模和管理，也可以作为产品协同设计、产品设计中的冲突协调、产品数据管理与流程控制的支撑系统。在这一应用中，需要增强 workflow 对产品数据及其相关集成文档的描述能力，并且需要在 workflow 技术中融入 CSCW 的技术和方法。

#### 1.1 workflow 的定义

workflow 管理联盟将 workflow 定义为：业务过程的全部或部分自动化，在此过程中，按照一系列过程规则，文档、信息或任务能够在不同的执行者之间进行传递与执行。

workflow 管理的一个主要目的就是将过程逻辑与嵌入在应用的任务逻辑相分离，这种分离的特性将使流程逻辑和任务逻辑能够被独立的修改，相同的任务逻辑能够被重用在不同的过程逻辑当中，因而达到了软

件重用的目的，从而更容易地集成异构、孤立的应用。应该提到，WfMC 的目的并不是自动化所有的任务。一些任务也设计了人的参与。甚至对一些自动任务来说，何时启动任务、何时完成任务和结束任务也可能取决于用户。

workflow 的重点是跟踪任务的状态、描述任务的先决条件以确定任务何时开始执行（任务间依赖）以及任务间的信息流。workflow 管理系统的基本思想是在正确的时间，以正确的顺序，将正确的任务分配给正确的人，并对业务全程进行完全地控制，实现对业务流程可控、在控的闭环管理。业务过程引入 workflow 管理系统的意义在于使工作有序，实现跨部门、跨应用系统的协调，协调各部门各岗位协同工作。workflow 管理系统的引入使高效协同工作已经成为可能。高效协同工作必须建立在基于计算机网络环境之上，以企业的业务过程为核心，并使财务、人员、物流等联系在一起，形成一个有机的整体，打破部门和职能藩篱，实现业务过程的高效、一致、全过程控制和信息共享，实现团队内的高效协同工作，适应管理不断变化的现实和持续完善的要求。

#### 1.2 workflow 基本术语

以下简要介绍一下 workflow 和 workflow 管理系统的一些基本概念，基本术语之间的关系见图 39.3-1。

1) 业务过程 (Business Process) 它就是工作从开始到完成的过程。workflow 由流程逻辑和路线规则组成。流程逻辑定义了任务的顺序和必须遵循的路线规则，还有截止期限以及由 workflow 引擎实现的其他业务规则。

2) workflow 管理系统 (Workflow Management System) 一个软件应用程序，它定义、管理并执行 workflow，计算机表示的 workflow 逻辑确定了该软件的执行顺序。workflow 引擎是运行时执行模块。

3) 过程 (Process) 过程是业务过程的规范视图，由一系列为共同目标联接在一起的协同的流程活动组成。

4) 活动 (Activity) 实现流程逻辑的工作任务的描述，一般分为手工操作和自动处理两类。活动是流程之中可被工作机调度的最小工作单元，要求有人或机器的参与

5) 过程/活动实例 (Process/Activity Instances)



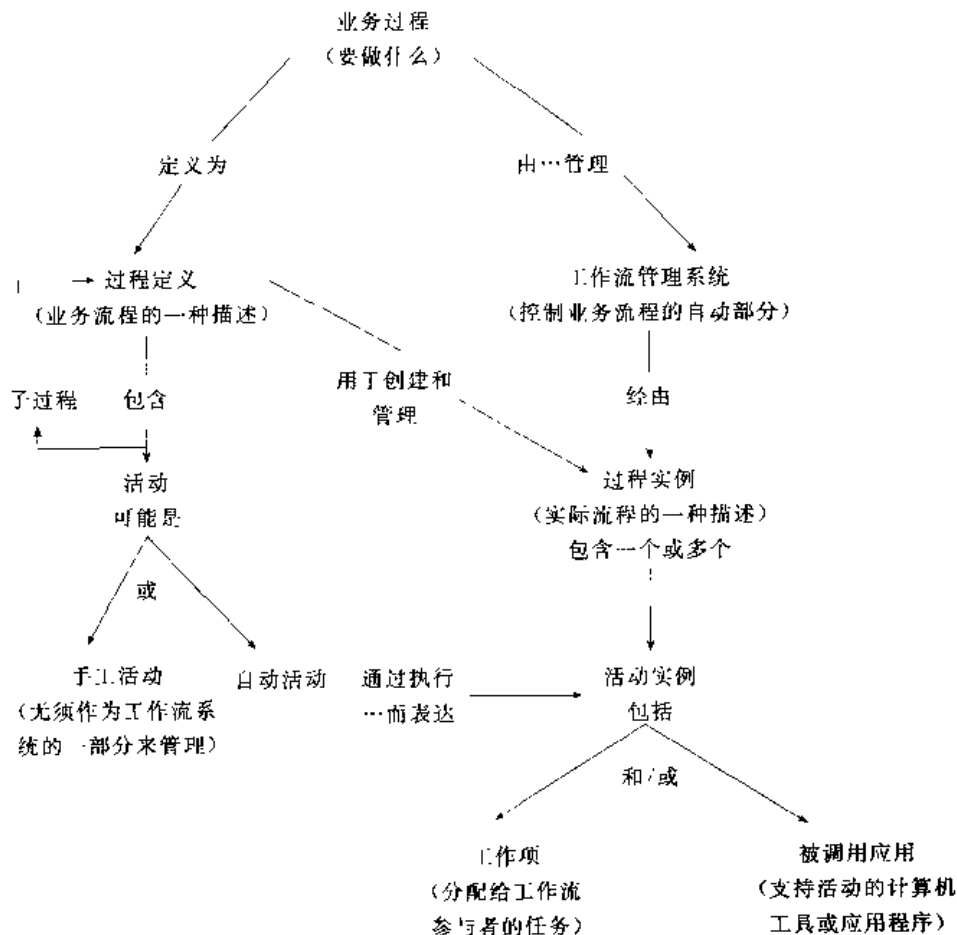


图 39.3-1 workflow 基本术语之间的关系

实际运行中的一个过程或活动。每个实例代表一个能独立控制执行、具有内部状态的线程，可被外界通过标识进行存取。

6) 过程实例由 workflow 管理系统创建、管理、终止，与过程定义一致。每个过程实例表示流程的一次单独执行，使用自己的过程实例数据，通常可以有自己独立的控制和审计过程。

7) 每个活动实例代表一个活动的一次调用，只和一个过程实例相关，并使用过程实例的数据。一个过程实例可能由若干活动实例组成，但一个活动实例不能同时属于几个过程实例。

8) 参与者 (Participant) 以下类型之一：资源集、特定资源、组织单元、角色（一个人在组织内部的作用）、人或系统（自动代理）。

9) 活动实例 (Activity) 组成过程定义中的一个逻辑步骤的任务。可以是自动的或人工的。自动指在过程操作过程中定义脚本和触发器的能力。过程定

义中的特定活动可以作为无人参与的任务来运行，自动化可以在手工或人力驱动的任务中执行业务规则。常见的一种自动活动就是截止期限管理，如果某个工作项在预定的截止期限之前未能完成，该管理可以自动发送一条提醒消息或触发一个延期程序。

10) 工作项 (Work Item) 代表过程实例中活动的参与者将要执行的工作。

11) 工作列表 (Worklist) 给定的 workflow 参与者或参与组的工作项的列表。工作列表是 workflow 引擎和工作列表处理器的接口的一部分。过程与工作列表的关系见图 39.3-2。

12) 工作列表处理器 (Worklist Handler) 管理用户或用户组与 workflow 引擎维护的工作列表之间交互作用的软件组件。它将工作项由 workflow 管理系统传递到用户，并将工作状态条件（如完成）在用户和 workflow 管理系统间传输。

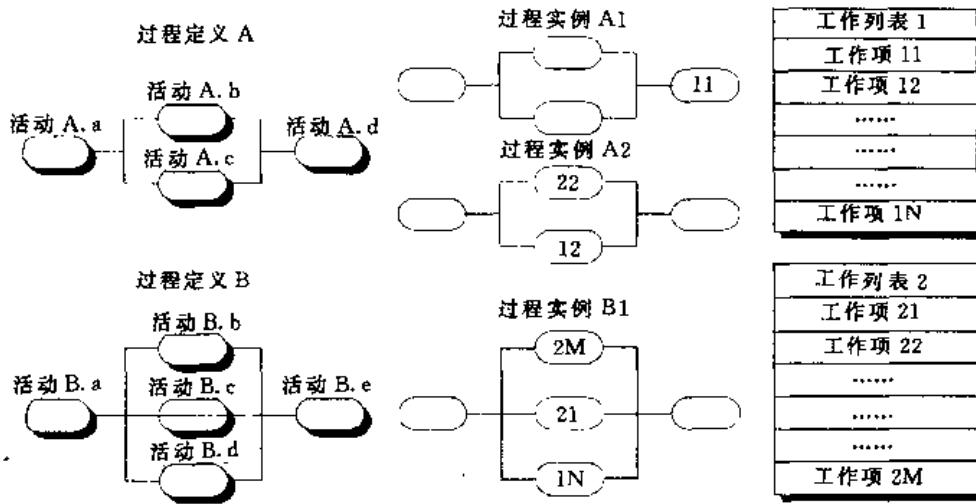


图 39.3-2 过程和工列表关系示意

### 1.3 workflow 管理系统类型

根据所实现的业务过程，workflow 管理系统可分为四类：

1) 管理型 workflow (administrative workflow) 在这类 workflow 中活动可以预定义并且有一套简单的任务协调规则，例如，大学里的课程选修，完成论文后的学位申请等。

2) 不确定型 workflow (ad hoc workflow) 与管理型 workflow 相似，但一般用来处理异常或发生机会比较小的情况，有时甚至是只出现一次的情况，这与参与的用户有关。

3) 协作型 workflow (collaborative workflow) 参与者和协作的次数较多。在一个步骤上可能反复发生几次直到得到某种结果，甚至可能返回到前一阶段。

4) 生产型 workflow (production workflow) 实现重要的业务过程的 workflow，特别是与业务组织的功能直接相关的 workflow。与管理型 workflow 相比，生产型 workflow 一般应用在大规模、复杂的和异构的环境下，整个过程会涉及许多人和不同的组织。

根据底层实现技术，又可将 workflow 产品分为三类：

1) 以通信为中心 以电子邮件为底层的通信机制。这种类型的 workflow 管理系统适合于协作型 workflow 和不确定型 workflow，而不适于生产型 workflow。

2) 以文档为中心 基于文档路由，它同外界应用的交互能力有限。许多基于表的管理型 workflow 可以用以文档为中心的工作流实现。

3) 以过程为中心 这种 workflow 系统对应生产型 workflow。它们一般建立在数据库之上，有自己专用的通信机制并且提供了同外部进行交互的接口。

所有的工作流管理系统一般都以提供以下三个主

要功能为特征 (图 39.3-3)：

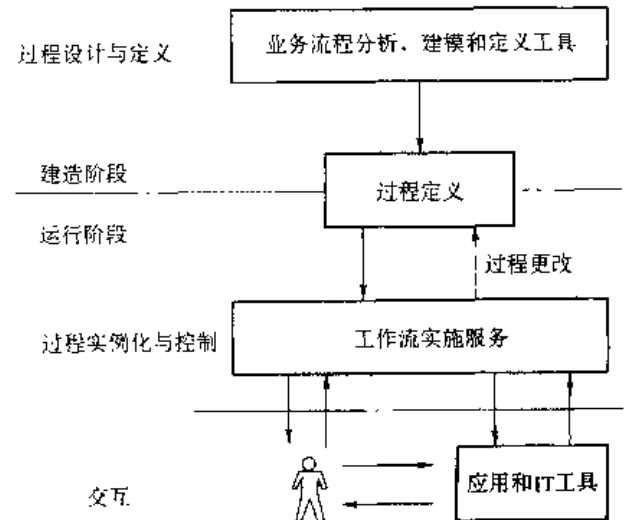


图 39.3-3 workflow 系统特征

1) 构建阶段的功能 对 workflow 过程以及组成它的活动进行定义和建模。

2) 运行阶段的控制功能 对 workflow 的运行进行管理，以及对每个 workflow 过程中的活动进行排序。

3) 运行阶段与用户和 IT 应用工具之间的交互作用。

与之相对应，workflow 一般都含有三个基本阶段：模型建立阶段、模型实例化阶段和模型执行阶段。在模型建立阶段，通过 workflow 建模工具，完成企业经营过程模型的建立，将企业的实际经营过程转化为计算机可处理的工作流模型。模型实例化阶段，给每个过程设定运行所需的参数，并为每个活动分配所需要的资源。模型执行阶段，完成经营过程的执行，主要是完成人机交互和应用的执行。

## 2 工作流管理系统参考模型

工作流管理联盟定义的工作流系统标准中包括一个参考模型及其 5 个接口的规范 (图 39.3-4), 这些规范确定了开发工作流产品所必须遵循的行业标准, 只有遵循这些规范开发的产品才可称为真正的工作流产品。

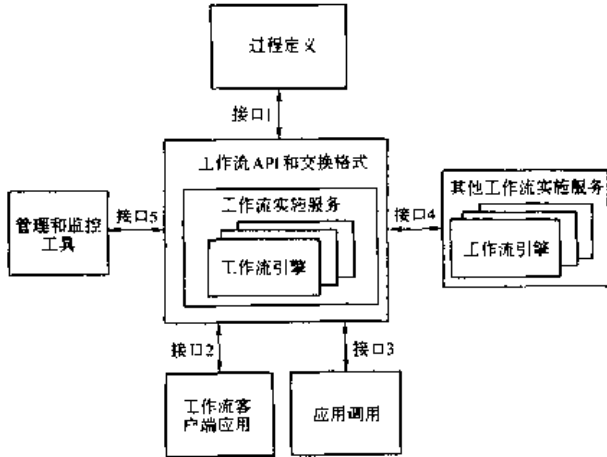


图 39.3-4 WfMC 的工作流系统参考模型

### 2.1 过程定义

过程定义 (Process Definition) 是业务流程的形式化描述, 用来支持系统建模和运行过程的自动化。过程可分解为一系列子过程和活动, 其定义主要包括描述过程起始、终止的活动关系网络以及一些关于个体行为的信息, 如组织成员、与 IT 相关的应用和数据等。在运行时期过程定义可以被工作流机解释 (图 39.3-5)。

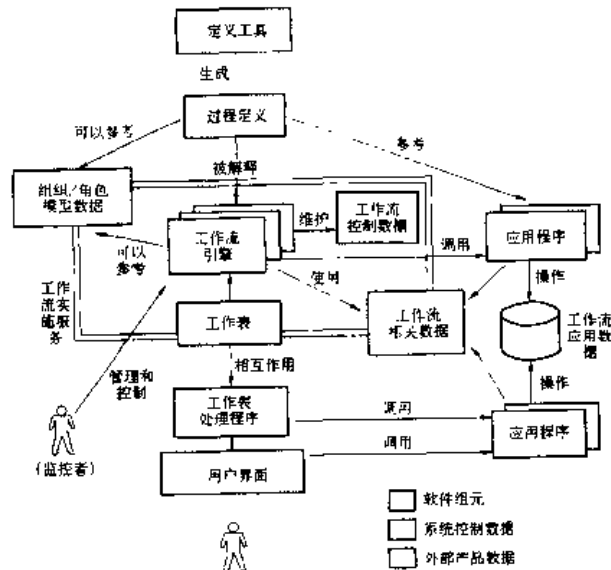


图 39.3-5 工作流产品的一般结构

#### 2.1.1 过程定义工具

过程定义工具是一个用来创建和更改过程定义的软件工具。该工具可以是一个业务过程管理软件的组

件、一个独立的应用程序或者一个工作流管理系统的组件。过程定义工具提供了重用已存储工作流元素甚至所有子流程的能力, 这使工作流应用程序开发者生产力更高。工作流模型不关心这些工具的特性, 也不关心在过程建立时期他们是如何交互的。这些工具可以作为工作流产品的一部分来提供, 或者一个单独的产品, 例如 BPR 工具集。

有的工作流产品提供了其自己的过程定义工具, 从而过程定义一般是保留在工作流产品范围内的, 并且可能或者不能被读/写信息的编程接口所访问。而使用单独的过程定义和执行服务器产品, 过程定义能够在不同的产品间进行转换, 并可以被其他产品访问。

过程分析工具建模工具和定义工具, 都要有在一个组织结构中模拟过程的能力 (尽管这不是工作流参考模型规定必须有的)。如果组织模型集成到了这些工具中, 那么过程定义将包含组织相关对象, 例如角色。这些都是与系统相关的控制数据, 例如角色、活动者间的关系, 可能会在过程执行期间被引用。

#### 2.1.2 工作流定义转换 (接口 1)

在建模或定义工具与运行时期工作流管理软件间的接口, 被称为过程定义导入/导出接口 (图 39.3-6)。这个接口的特点是: 转换格式和 API 调用, 从而支持过程定义信息间的互相转换。这个接口也支持已完成的过程定义间的互相转换。或过程定义的一部分。例如, 过程定义的改变或者活动中属性的改变。

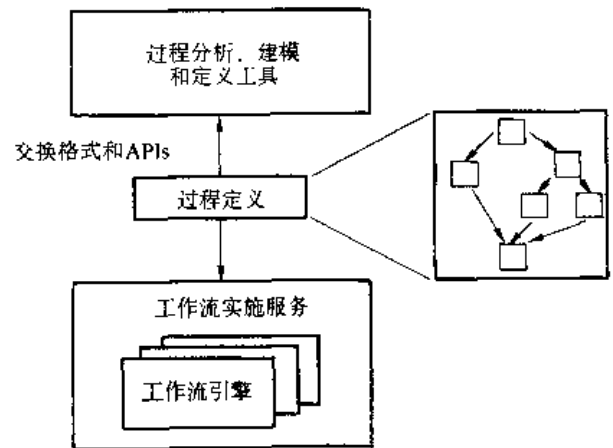


图 39.3-6 过程定义接口

首先, 把建立阶段与运行时期环境进行了分离, 可以使用一个建模工具来产生过程定义, 这个过程定义可以作为很多个不同工作流运行时期产品的输入。从而用户可以单独地选择建模工具和工作流运行时期产品。

其次, 可以为几个工作流机输出过程定义, 这几个工作流机合作未构成分布式的工作流执行服务器。

WFMC 在此部分作了以下两个方面的工作:

1) 提出了一个元模型,可以用来表示过程定义中的对象,对象间的关系和属性。这个元模型为不同的产品间的过程定义相互转换奠定了基础,并形成了一套转换格式。

2) workflow 系统间或 workflow 系统与过程定义产品间的 API 调用,提供了公共的方法来访问 workflow 过程定义。访问可能是读、读/写或者只写操作,并且操作标准对象集合(在元模型中定义的对象集合)或者产品自己的对象集合。

### 2.1.3 基本元模型 (A Basic Meta-Model)

WFMC 开发了一个过程定义的元模型(图 39.3-7)。元模型中定义了基本的对象类型集,来满足简单的过程定义相互转换或者由开发者具体扩展,或者在增加的功能中定义另外的一些属性来增加更多的对象类型。在分布式 workflow 服务器中,可能要在过程定义时,为每个 workflow 机分配活动,可以作为活动的一个附加属性。过程定义能影响安全性与管理。

定义的交换格式,要支持符号命名方案,这些符号可以映射到 workflow 执行服务器中的实际名称与地址。这种映射可以使用动态地址定位机制来实现(例如目录服务器),也可以使用其他的外部过程定义机制实现。也有其他的一些行业在相关的方面作研究,例如过程建模和 CASE 转换工具,WFMC 提出的方法也适用于其他行业,预先定义适当的转换格式。

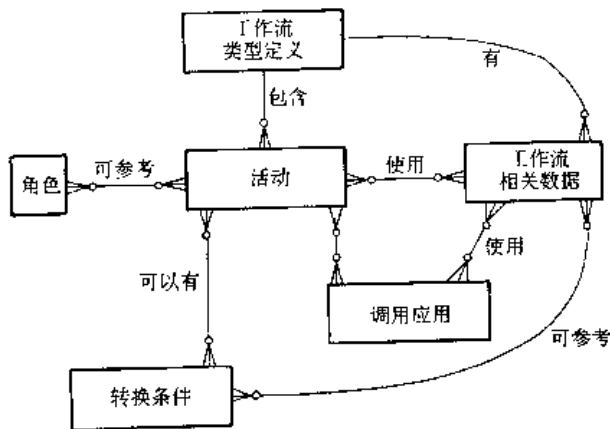


图 39.3-7 基本过程定义元模型

## 2.2 workflow 客户端功能

### 2.2.1 workflow 客户端应用程序 (Workflow Client Applications)

任务表处理器是在需要调用人类资源的活动中,用来与终端用户进行交互的软件。任务表处理器可以作为 workflow 产品的一部分提供给用户,也可以由用户

自己开发。在其他情况中, workflow 可能要与普通的办公系统进行集成,例如 Email, 来为终端用户提供一个统一的任务管理系统。这就要求在 workflow 执行服务器与 workflow 客户端应用程序间有一个非常灵活的通信机制,来构建各种可能遇到的运行系统。

在 workflow 模型中,通过客户端应用程序与 workflow 机间的定义良好的接口进行交互。在这个接口中包含任务表——由 workflow 机分配给用户的任务序列。最简单的情况是, workflow 机访问任务表,来把任务分配给用户;任务表处理器访问任务表,向任务表中添加任务项。有许多不同的产品来实现任务表的交互。任务表中任务项的激活(例如,自动应用程序,连接 workflow 相关数据),可能是由 workflow 客户端应用程序或者终端用户控制的。在 workflow 客户端应用程序与 workflow 执行服务器间定义了一系列的方法,用来向任务表中添加任务项、从任务表中删除完成的活功激活临时挂起的活动等。

任务表处理器也可以调用应用程序,或者直接调用,或者由终端用户调用。通常希望,任务表处理器的应用程序调用范围能够受到运行环境的限制,尽管这样会给模型带来通用性的限制,但这种情况是一直存在的。

与任务表相关的部分活动的数据,是任务表处理器用来调用应用程序所必须的信息。当应用程序数据是强类型时,在任务表处理器中要存放一个联接,用来实现程序的调用。在其他情况中,在任务表处理器与 workflow 机间要进行完全的应用程序名称和地址信息的交换;这时, workflow 客户端应用程序也可能实现一些应用程序调用接口(接口 3)中的功能,来获得必要的信息。

任务表中可能要包含一个过程中的几个不同实例的相关任务,或者包含几个不同过程中的一个共同活动项。一个任务表处理器可能要与几个不同的 workflow 机,几个不同的 workflow 执行服务器进行交互(按照每个产品的实现,为每个过程单独维护一个物理上分开的任务表,或者任务表处理器把几个不同的任务表联合到一起,呈现给终端用户)。

因此,客户端 workflow 应用程序与 workflow 机间的接口必须十分灵活,来满足下面的几方面功能实现的多样性:

- 1) 过程和活动表示符。
- 2) 资源名和地址。
- 3) 数据引用和数据结构。
- 4) 可选择的通信机制。

### 2.2.2 workflow 客户端应用程序接口 (接口 2)

满足上述需求的方法,在标准 API 集后,可以为从 workflow 应用程序到 workflow 机和任务表的访问提供一

致的形式,而不管产品的实现特性。API 与其参数可以映射到几个不同的通信机制上,来适应各种不同的 workflow 实现模型,如图 39.3-8 所示。

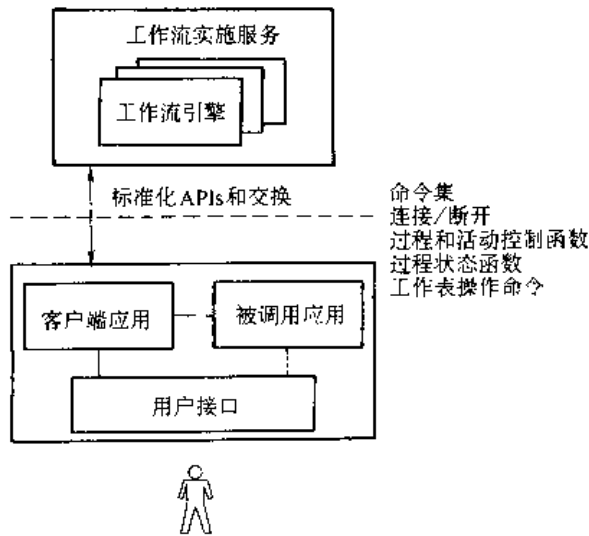


图 39.3-8 客户端应用接口

## 2.3 应用程序调用功能

### 2.3.1 应用程序调用

所有的 WfMC 产品都没有足够的逻辑单元,知道如何调用所有的应用程序,这些应用程序存在异种的产品环境中。这就需要,能够处理在所有平台下和网络环境中进行调用的逻辑,并需要能使用公共格式和编码进行应用数据或相关数据传递的方法。

然而,许多 workflow 系统能够使用更多受限制的应用程序,特别是那些采用强制数据类型和直接与应用程序相连的系统。在其他情况中,应用程序对操作的调用,可能是通过标准的交换机制来实现的,例如 OSI TP 协议或者 X.400。一些实现使用了“应用程序代理(Application Agent)”把这些在标准接口之后的各种方法包含在 workflow 执行服务器中。也有可能开发“workflow enabled”应用工具。这种工具使用标准的 API 集来与 workflow 执行服务器进行通信。来接收应用程序数据信号和响应活动事件,等这些在标准接口之后被应用工具直接调用,也可以被应用程序代理过程调用,作为与其他应用程序(不包含任何 workflow 技术的程序)交互的前端。

### 2.3.2 应用程序调用接口(接口 3)

调用应用接口如图 39.3-9 所示。在简单的情况中,workflow 机在本地处理应用程序调用,使用过程定义中的信息来确定活动的性质、将要调用的应用程序的类型和所需的数据。被调用的应用程序可能存储在工

作流机中,或者与 workflow 机一同存储在相同的平台下,或者存放在一个独立的网络访问的平台中;过程定义中有足够的应用程序类型和寻址信息(workflow 机的特殊需求),来实现应用程序调用。在这种情况下,应用程序命名与寻址的协定是处于 workflow 机与过程定义之间的。

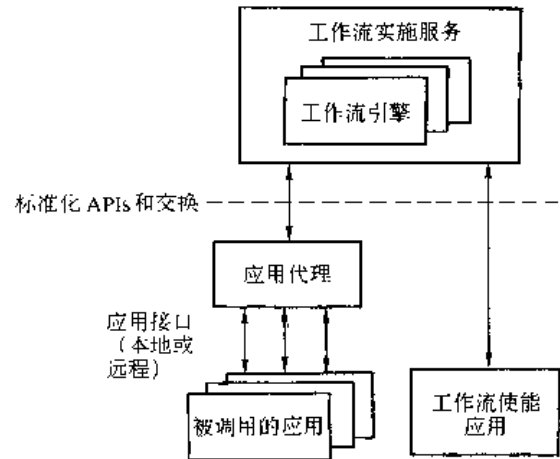


图 39.3-9 调用应用接口

应用程序调用 API 的详细语法、语义作为 WfMC 规范的一部分给出。操作覆盖了一些不同的基本接口,包括上表中的一部分,其中一些操作是同步的,一些是异步的。API 的操作可以是单线程的,也可以是多线程的。后者使用活动 ID 来区分线程。下边是应用程序调用可以使用的一些命令概括:创建会话、活动管理功能、数据处理功能。

更复杂的情况,如异种 workflow 机间的协同工作,可能需要在 workflow 机间传递应用程序调用信息,或者作为运行时期数据交换的一部分,或者通过在过程定义阶段后导入过程定义来实现。

## 2.4 workflow 互操作能力

WfMC 的一个主要目标是提供一个标准,使任务能在不同开发商的 workflow 系统产品之间进行无缝传递。

workflow 产品的特性十分多样化。在 WfMC 的协同工作标准中,没有强迫开发商必须在提供一个只面向用户需求的产品和只考虑协同工作的,二者中作选择。

WfMC 把焦点聚集到开发多种不同的协同工作框架,这些框架可以操作一系列标准的协同工作,从简单的任务传递到整个 workflow 系统的协同工作(包括过程定义转换、workflow 相关数据交换、通用的界面等)。简单的协同工作,WfMC 的协同工作定义将在最初就能支持;而复杂的协同工作,还需要进一步的研究。

### 2.4.1 协同工作模式

尽管可以开发一个非常复杂的协同工作框架,由

许多个工作流机构成一个执行服务器，但是这种框架不会在近期实现，因为这需要所有的工作流机都可以解释一个公共的过程定义和共享公共的工作流控制数据集，事实上是维护异种工作流机间的一个共享过程视图。现阶段更现实的目标是能够在运行时期传过程的某些部分，来支持不同的执行服务器运行。

WFMC 定义了 4 个协同工作模式，包含多种协同工作能力级别。

**模式 1——链接 (Chained)**

如图 39.3-10 所示，这个模式中过程 A 中的一个连接点，连接到过程 B 中的一个连接点。尽管在图中给出了过程的开始连接点与结束连接点，但这并不仅是为了图

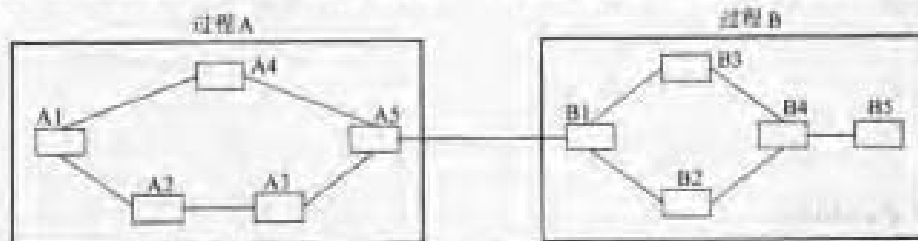


图 39.3-10 链接模式

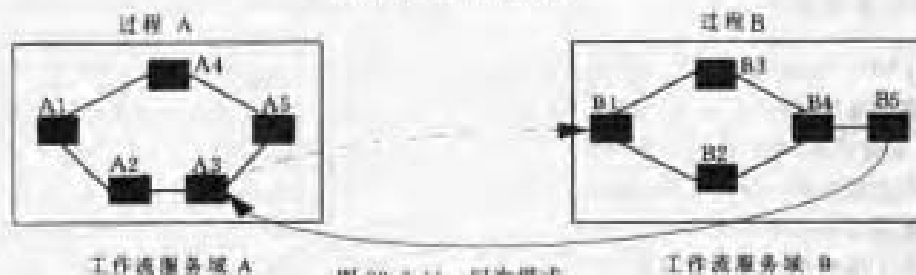


图 39.3-11 层次模式

在图 39.3-11 中，工作流执行服务器 A 有一个活动 (A3) 会在过程 B 结束后执行，过程 B 在工作流执行服务器 B 中执行，并在过程 B 执行结束时，执行服务器会把控制返回给执行服务器 A。活动控制的传递，要经过两个工作流执行服务器间的应用程序网关功能或者直接的 API 调用。图中描述的是一种简单情况。

**模式 3——P2P (Peer-to-Peer)**

这种模式存在于混合环境。图 39.3-12 中给出了一个复合过程 C，C 中的活动需要在多工作流机执行服务器中执行，形成一个共享范围，活动 C1、C2 和 C5 由执行服务器 A 执行，活动 C3、C4 和 C6 由执行服务器 B 执行。在这个模式中，过程是完全由任务到任务来推进的，不需要用户或管理员参与，交互只在工作流引擎间存在。

在此模式中，需要两个工作流执行服务器都必须支持公共通信 API 集，并且两者都能解释一个公共的过程定义，或者从一个公共的建立过程引入到这两个

解的目的。连接点可以是过程中的任何一个活动。

这个模式支持在两个工作流环境中传递一个任务，这个任务会在第二个环境中独立执行，不需要同步。可以使用一个网关应用功能来实现这种模型，网关负责处理数据格式转换过程和名称映射等，例如在两个执行服务器中使用标准 API 调用。

**模式 2——层次式 (Nested Subprocesses)**

一个过程可以在全部在一个工作机范围内执行，封装成父过程的一个任务，在一个与执行父过程不同执行服务器中运行。在父过程与封装过程间存在一个等级关系，构成父过程的子过程。这种等级关系可以延伸到很多层，形成嵌套子过程。产品可以实现回归，也可以不实现。

执行环境中，或者在运行时期从两个服务器中导出，工作流相关和应用程序数据，也要能在多个异种工作流引擎间进行传递。

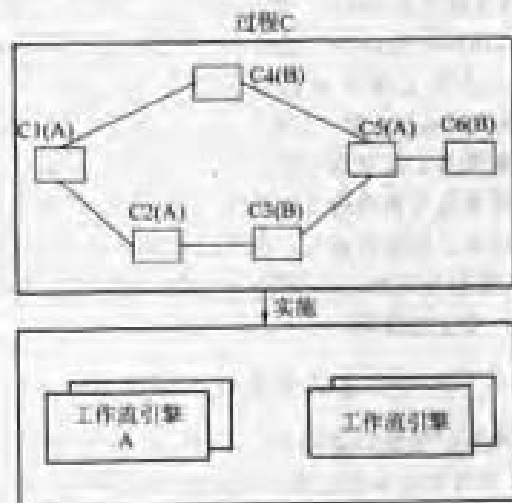


图 39.3-12 P2P 模式

#### 模式4——并行同步 (Parallel Synchronized)

如图 39.3-13 所示,在这个模型中,允许在本质独立执行两个过程,可以访问独立的执行服务器,但是在两个过程间要有同步点。同步需要一旦每个过程都达到在其执行序列中的一个预先定义的点,就要激发一个公共事件。这种机制可以用来在交叉执行的线程中进行过程时间安排,检查恢复数据,在不同的过程实例

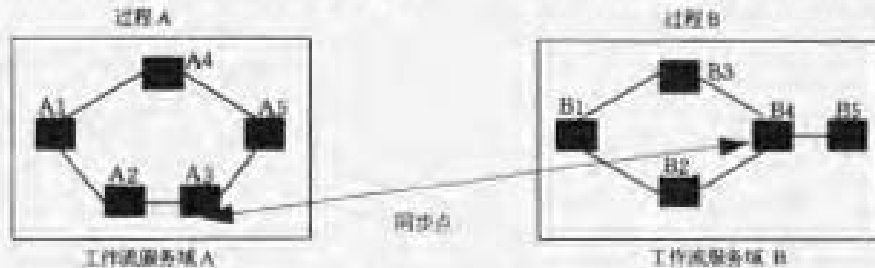


图 39.3-13 并行同步模式

件性分为八个级别。

1) 无互操作性 属于这个级别的工作流系统,没有提供与其他工作流系统通信的手段,因而不具有互操作的能力。

2) 共存 属于这个级别的工作流系统,不具备和其他工作流系统互操作的功能,但可以和其他工作流系统同时运行在同一个运行环境中。属于这个级别的工作流系统,其目的不是提供系统之间的互操作,而是为了增加可运行在同一平台上的工作流系统的数量,使一个完整的工作流,划分为几个部分,在不同的工作流系统中完成,相应地也扩大了用户选择适合自己的工作流系统的范围。在这种级别下,工作流系统之间的接口由人工提供(如果共存的工作流系统都遵循 WFMC 标准开发,也可通过标准接口通信)。

3) 网关 工作流网关是允许在不同工作流系统之间传递实体(过程、活动或者是工作项等)的一种体制。在需要数据格式转换时,工作流网关也提供工作流系统之间的数据映射功能,实现工作流系统之间的通信。在存在多个工作流实例的环境中,网关还具有“数据路由”的功能。具体实现时,工作流网关可以是工作流系统的一部分,也可以是单独的应用程序。WFMC 定义了两种工作流网关:单独网关和公共网关 API。

4) 有限公共 API 子集 有限公共 API 子集是工作流系统共同定义一组核心 API。工作流系统通过这组核心 API 进行交互。这种级别的互操作不仅要定义一组公共 API,还必须定义工作流系统之间数据交换的标准格式。对应用来说,API 可以是标准的接口,也可以是数据封装的形式,标准的数据交换格式则一般通过定义“过渡数据交换格式”的形式实现。

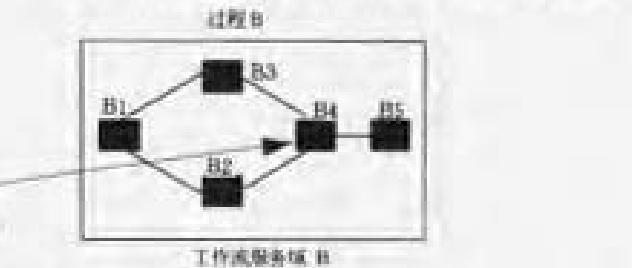
5) 完整工作流 API 完整工作流 API 是定义工作流系统所有操作的 API 规范,工作流系统之间的交

互传递工作流相关数据等。

一对匹配的任务,可以在每个过程中的某个点同步,这需要一个协调和跟踪机制事件,另外两个执行服务器都要能够从连个过程定义中识别任务。

#### 2.4.2 工作流互操作分类

按照国际工作流管理联盟的定义,工作流的互操



互通过 API 来进行。与有限公共 API 接口相比,完整工作流 API 级别的互操作包含的工作流互操作范围要广泛的多,而且这些 API 是在更高的层次上抽象出来的。当然,完整工作流 API 的实现与有限公共 API 一样,并可以认为它是有限公共 API 的扩展。

6) 共享定义格式 共享定义格式是工作流系统之间的所有过程定义(包括过程属性、信息路由描述、访问权限等)采用统一的标准。如 WFMC 制定的工作流过程定义语言(Workflow Process Definition Language, WPDL)就是这种级别的互操作。通过这种方式,过程定义在不同工作流系统之间共享,满足系统互操作性的要求。在实现形式上,这种级别必须支持导入(Import)和导出(Export)两个接口。

7) 协议兼容 协议兼容指对工作流系统客户和服务器之间的通信(包括过程定义的传递,工作流的事务处理和错误恢复等)都进行标准化。为了得到这个级别的互操作,工作流系统必须提供多种通信机制(常见的有 X.400、远程过程调用、其他协议),因此实现难度较大。

8) 等同式 等同式指工作流系统所有用户接口都完全相同(或至少感觉上相同)。由于商业和实际的原因,这种级别的互操作很难实现。上述八个级别的互操作性逐级提高,工作流系统实现难度也是依次不断增加。为了满足不同需求,工作流系统实现时可以选择不同的互操作级别。

#### 2.4.3 工作流互操作的实现技术

从实现方式来看,工作流互操作可有直接通信、信息传递、桥和共享数据存贮四种方式,这四种方式各有特点,使用范围也不尽相同。根据工作流参考模型,可采用工作流引擎之间直接通信和桥的方式。

(1) workflow 引擎直接通信

这种方式的工作流互操作实现形式实质上是在 workflow 引擎之间定义一组标准的工作流应用程序编程接口和交换格式 (Workflow APIs and Interchange

Formats, WAPI), 通过 workflow API 和数据交换格式, workflow 引擎可以直接通信, 从而实现 workflow 互操作性。接口 4 正是为实现 workflow 互操作而定义的, 如图 39.3-14 所示。

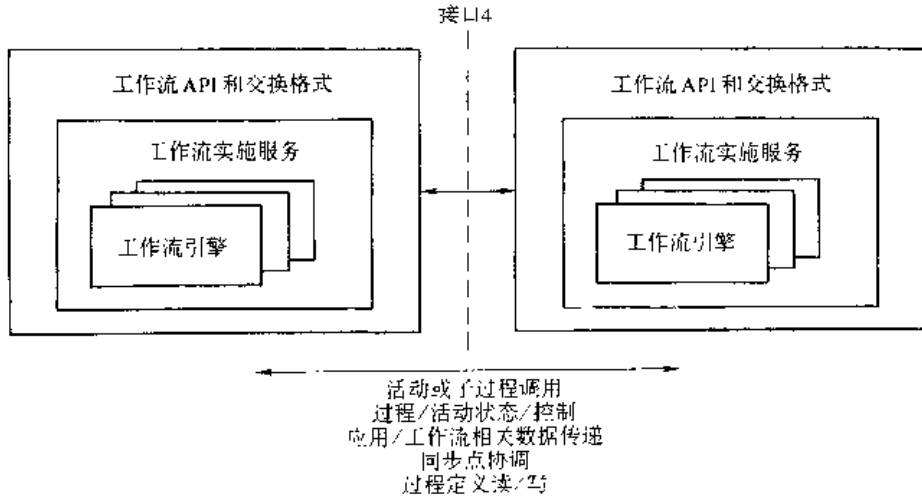


图 39.3-14 workflow 互操作接口

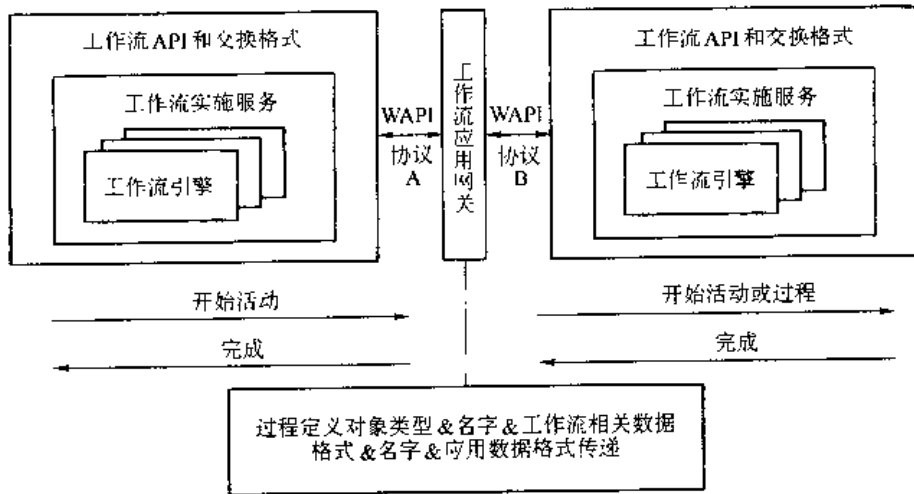


图 39.3-15 workflow 网关实现互操作

(2) 桥

采用“桥”来实现 workflow 的互操作, 实质就是由 workflow 网关来传递数据并对数据进行映射, 即到达互操作第三级别的互操作性要求, 这种实现形式如图 39.3-15 所示。

2.4.4 workflow 管理系统互操作性存在的问题

随着 workflow 技术迅速发展, workflow 管理系统互操作技术仍进展缓慢, 采用上述四种典型互操作设计模式只能部分解决 workflow 互操作问题。目前, workflow 互操作技术仍存在以下三个不足的方面:

1) workflow 模型缺乏描述互操作性的能力 虽然 WfMC 提出了一套 workflow 管理系统的标准, 也定义了

workflow 系统的参考模型, 但是该模型仅能通过 WAPI 来实现 workflow 互操作, 这显然不能完全描述和表达 workflow 系统的互操作。此外, workflow 参考模型及相关标准没有对 workflow 通信机制进行定义, 而在 workflow 系统实现时, 对选择何种通信协议存在很大的分歧, 很难实现真正意义上的互操作。

2) workflow 互操作过程中缺乏健壮性 workflow 实例在运行过程中可能出现资源访问冲突, 使得活动在数据操作上出现相互重叠。这种情况在 workflow 互操作过程中表现尤为明显。目前, 对互操作过程中 workflow 系统健壮性还很难保证。

3) 缺乏互操作情况下事务处理的能力 目前大部分 workflow 系统不支持事务的概念, 不能保证在工作



流系统之间互操作下一个活动执行的 ACID (Atomicity, Consistency, Isolation, Durability) 属性, 而且当 workflow 非正常中断时, 也无法保证正确的恢复数据。综上所述, 目前 workflow 管理系统以及 workflow 互操作的发展还有许多问题亟待解决, 而对这些问题的解决也正是今后 workflow 管理系统和 workflow 互操作技术的发展方向。

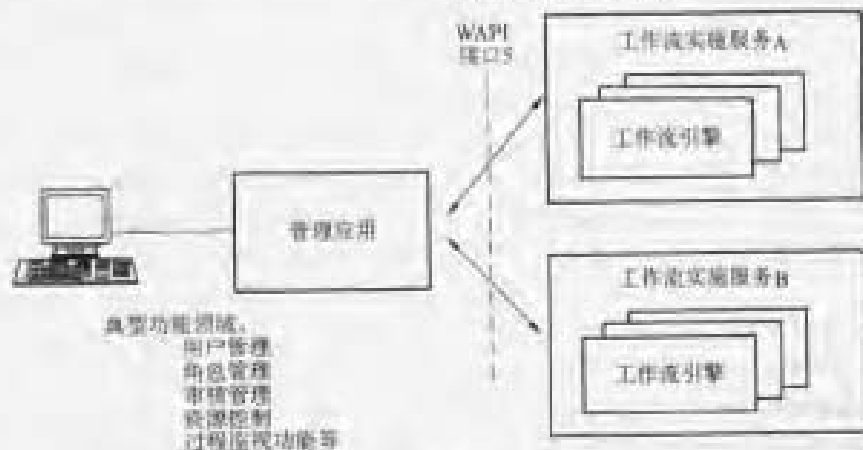


图 30-3-16 系统管理和监视接口

尽管, 过程状态命令在接口定义中已经描述了, 但一致认为, 在某些行业中需要进行全部状态监视和提取信息的功能。WFMC 提出的接口是要让用户能够得到 workflow 运行状态的完整视图, 无论是什么样的 workflow 系统, 同时, 他希望能提供一套全面的函数集, 进行系统管理, 包括安全性控制和权限。

接口中包含 WAPI 集中的一些具体命令, 来操作管理和监视功能。另外, 进一步的讨论, 期望能够确定在什么范围内, 这个接口可以使用现有的协议 (如 CMIP, SNMP), 来设置, 恢复管理状态和统计信息 (定义在开放 MIB 中——Management Information Base)。

## 2.6 工作流故障恢复和异常处理

工作流执行失败的情况可分为两类: 一类是因系统故障 (如掉电、断网) 导致工作流执行中止; 另一类是执行条件不成立或授权代理 (人或计算机) 的干预, 导致工作流执行中止或取消已提交的任务。当出现执行失败时, 工作流停留在不一致的状态上。为了将其恢复到正确的状态上, 则有两种手段, 一种是采用应急任务, 向前恢复; 另一种是采用补偿任务, 向后恢复。为此, 可利用持久性对象仓储服务, 为工作流的执行建立日志, 记录工作流当前的和历史的执行状态。此外, 在工作流执行过程中, 还可能出现其它异常情况, 并且大多数情况是不可能在设计工作流时预测的, 如数据冲突、人员变化、规则调整等。因此, 需要设置公告处理

## 2.5 系统管理

WFMC 规范的最后关注的是, 为管理和监视功能开发公共的接口标准。这样, 一个开发商的产品就可以用来管理其他 workflow 机的运行, 通过公共的接口几个不同的 workflow 执行服务器可以共享管理和监视功能, 如图 30-3-16 所示。

机制, 当出现系统不能处理的异常时, 一方面保存好用于以后恢复的所有状态, 另一方面报警通知操作人员进行人工干预。

## 2.7 工作流规范

在一个由 Web 服务构建而成的合作应用程序里, 应用程序中的业务过程是一组任务, 这些任务的参与者是 Web 服务, 而 workflow 控制极为重要, 完全不同的 workflow 之间的交互也不可避免。因此, 需要一个标准去描述公共流程, 组合、专用 workflow 和其他常见的工作流构件。尽管现在已经有了一些被提议的工作流标准, 但这种 workflow 互操作性还未被人们确定。以下是近年来产生的各种 workflow 规范和定义:

1) WI-XML——工作流管理联盟 (Workflow Management Coalition, WFMC) 中的 WI-XML 和工作流参考模型 (Workflow Reference Model)。WI-XML 是一种基于 XML 的工作流互操作性信息的编码。工作流参考模型是一种底层 workflow 系统体系结构的描述。目前 WI-XML 没有与 SOAP 和 WSDL 绑定。

2) WSFL——IBM Web 服务流语言 (IBM Web Services Flow Language)。指定了 Web 服务组合的两种类型: 一个被认为是流模型 (flowModel) 的可执行业务流程; 另一个被认为是统一模型 (globalModel) 的业务合作, 与 SOAP、UDDI 和 WSDL 兼容。

3) XLANG——Microsoft 的 XLANG: 用于 BizTalk 的业务模型语言, 该语言是可运行 EAI 的。

NET 组件。BizTalk 编制 (BizTalk Orchestration) 是工作流引擎, BizTalk 编制设计器 (BizTalk Orchestration Designer) 是基于 XLANG 的可视化业务流程模型工具。

4) BPEL4WS——用于 Web 服务的业务流程执行语言 (Business Process Execution Language for Web Services) 是用于 Web 服务编制、工作流和组合的 WSFL 和 XLANG 的协作合并。该语言还尚未被提交到 IT 标准组织。

5) EbXML——BPSS 电子商务过渡工作组 (eBusiness Transition Working Group) 继承了业务流程规范方案 (Business Process Specification Schema (BPSS)) 的 ebXML 层中的工作流对话和编制, ebXML 定义了许多基于 XML 的电子商务的协议和层。

6) WSCI——Sun/BEA/Intalio/SAP 联盟的 Web 服务编排接口 (Web Services Choreography Interface) “是一种基于 XML 的接口描述语言, 该语言描述了参与和其他服务的编排交互作用的 Web 服务所交换的信息流。”

7) WSCL——W3C 的 Web 服务对话语言 (W3C's Web Services Conversation Language); Hewlett-Packard 向 W3C 的提交, 该提交允许定义 Web 服务的抽象接口 (也就是 Web 服务支持的企业级对话或公共流程), 以及交换的 XML 文档及其文档的排序。

8) PIPs——RosettaNet 的伙伴接口流程 (Partner Interface Process): 定义了贸易伙伴与指定的系统到系统 (system-to-system) 的基于 XML 的对话之间的业务流程。许多 PIP 被用来定义各种伙伴情况。

9) JDF——CIP4 的工作定义格式 (Job Definition Format) 是一种即将使用的用于图形艺术 (Graphics Arts) 工业的工作流工业标准, 该标准用于简化不同应用程序和系统之间的信息交换。

### 3 工作流管理技术研究现状

#### 3.1 工作流建模技术的研究现状

目前工作流建模技术还远没有成熟, 处于百家争鸣的时代。对当前流程建模领域出现过的主流建模方法进行归纳分析, 可以把它们分成 5 大类: 脚本语言、基于网的方法、基于规则的方法、基于逻辑的方法和代数方法。

1) 脚本语言 (Script Language) 脚本语言中有描述控制流和数据流的语句结构, 它提供了一种简明的描述方法, 对于有经验的设计者来说更易于使用。它的主要缺点是缺乏流程本身的形式化语义, 语言的语

义主要是供语言解释器使用。

2) 基于网的方法 适合于流程的可视化建模, 一般都使用状态变迁网, 流程中的活动用结点表示, 控制流用边表示。使用状态变迁网的主要问题在于是否有形式化的语义, 大多数工作流产品的可视化建模方法都缺乏形式化语义。在具有形式化语义的状态变迁网中, 使用得最多的就是 Petri 网和状态图 (State Charts)。具有形式化语义的基于网的方法可以转换成其他建模方法, 如基于规则的、时序逻辑的和脚本语言的方法等。

3) 基于规则的方法 目前研究得最多的基于规则的方法是 ECA (Event/Condition/Action, 简称 ECA) 规则, ECA 规则最早用于 AOODBS, 而后又被用于工作流管理领域。ECA 规则具有形式化基础, 也可以转换成其它的建模方法, 如 Petri 网、时序逻辑等, 但是 ECA 规则的可视化工作量比较大。另外, 如果规则集比较大时就难以管理。

1) 基于逻辑的方法 基于逻辑的方法适合于描述系统的动态性, 其中时序逻辑是一种常用的方法, 它具有很好的形式化基础, 验证工作流模型的属性比较方便。但是时序逻辑的主要缺点是很难实现可视化, 不容易转换成其它的描述方法, 描述业务流程的系统行为太复杂。

5) 基于代数的方法 过程代数 (Process Algebra) 主要还是局限在理论探讨上, 在工作流管理领域用得很少, 只有一种基于过程代数的描述语言 LOTOS 被用于工作流管理领域。代数方法的主要缺点类似于基于逻辑的方法, 并且在自动执行和形式化验证方面表现得比基于逻辑的方法更差, 建模方法缺乏直观性, 难以理解。

根据以上的比较可以看出, 各种建模方法各有优缺点。但从总体上来看, 脚本语言、基于网的方法和基于规则的方法更具有吸引力。

#### 3.2 工作流管理系统的研究现状

在工作流技术研究中, 比较著名的有 IBM 公司 Almaden 研究中心的 Exotica、佐治亚大学计算机系的 Meteor、WIDE 以及 Mentor 等研究项目。

(1) 基于持久消息队列的分布式工作流管理系统——Exotica

Exotica 的工作流管理系统由许多具有自治能力的节点组成。每个节点的运行都独立于其他节点, 节点之间通过可靠的消息队列通知对方, 过程中某一步已经完成。这种运行机制避免了节点在过程运行中不断与服务器通信所造成的瓶颈, 增强了系统的鲁棒性, 如果其中一个节点出现故障停止运行, 其余节点仍能正

常运转。

Exotica 使用 IBM 本身的工作流产品 FlowMark 提供的建模工具建立模型。为了更好地处理活动和应用之间的信息传递,该建模工具引入了输入容器、输出容器和数据流等模型元素。数据流由过程模型中的数据连接弧组成,它描述了各个输入容器和输出容器之间的映射关系,使得活动间可以 Exotica 面向分布式环境下的用户,其过程实例的运行分布在整个运行环境内,由各个节点根据模型的定义和具体实例运行情况推进实例的运行。

#### (2) 具有自适应能力的工作流管理系统

Meteor

美国佐治亚大学研制的 Meteor 系统是一个支持多范型的工作流管理系统,主要用于处理医疗保健应用。多范型是指该系统能够支持分布异构环境下的企业内和企业间的各种工作流。这些工作流可以是数据库管理系统和分布式事务处理系统中的事务,也可以是 EDI 等特殊应用。该系统采用了完全分布的体系结构,并实现了对工作流任务的分布调度。Meteor 还提供了一个自动的代码生成器,能够自动地将图形化的工作流模型转换为实际运行代码,并能对模型进行动态修改,可有效地支持企业实现动态经营过程重组。Meteor 系统的体系结构由建模工具、自动代码生成器、运行环境、监控器和数据库组成。

Meteor 提供了一套图形化建模工具,包括流程设计器、数据设计器和任务设计器。其中,流程设计器用于定义整个工作流模型中各个活动间的关系;数据设计器用于定义执行活动所使用和传递的数据;任务设计器可支持五种不同类型的活动,即非事物型、事物型、Web 型、人机交互型和两阶段提交型,并描述了如何激活这五种不同类型的活动。Meteor 建立的模型将以工作流中介语言 (Workflow Intermediate Language, WIL) 的形式保存起来。WIL 能够记录活动间的前驱和后继关系、活动间所传递的数据对象、数据对象的定义以及活动的具体描述和激活方法等等。Meteor 为基于 CORBA 的运行系统和基于 Web 的运行系统分别提供了相应的代码生成器,它以 WIL 格式的工作流模型为输入,为活动管理器提供实际运行代码,包括调度、活动激活、数据对象获取以及出错恢复等功能的代码,它还可以输出代码,对数据设计器定义的数据对象进行维护和操作。Meteor 提供的工作流运行支持系统有 ORB Work (基于 CORBA 的完全分布式工作流执行系统)、NeoWork (基于 CORBA 的集中式执行系统) 和 WebWork (基于 Web 技术的完全分布的执行系统)。Meteor 还对各种异常情况设计了专门机制和函数,负责检测异常情况。当异常发生时,可以

交给异常处理函数处理,也可以执行一个补偿活动来取消出错活动所产生的影响。

#### (3) 基于分布式主动数据库技术的工作流管理系统——WIDE

WIDE 是由西班牙、意大利和荷兰等国的五个合作单位协同开发的工作流管理系统,主要目的是利用分布式数据库和主动数据库技术来实现工作流管理,并提供先进的、面向应用的软件产品。WIDE 工作流模型包括:组织模型、信息模型和过程模型。WIDE 是对 WfMC 参考模型的一种扩展。它不仅定义了工作流的基本要素,而且还支持组织模型建模、复杂的活动约束分配、动态流程控制、复杂过程结构以及工作流事务处理。

1) 组织模型记录了企业内组织结构和资源的信息。

2) 信息模型定义和维护工作流实例运行过程中所需的全部数据,并管理这些数据的使用范围和表示方式。

3) 过程模型是 WIDE 工作流模型的主模型。它的活动具有以下属性:前条件(规定活动开始的标志)、活动所要进行的操作、后条件(表明活动结束的条件)、角色限制以及对系统定义和用户自己定义的异常情况的处理。

WIDE 提出的工作流管理系统包括三层:最底层是数据库层。第二层为基本访问层 (Basic Access Layer, BAL),实现对数据库的屏蔽,BAL 为用户提供了一个面向对象数据库的操作接口,并将其映射为底层的关系型数据库操作接口。最顶层是服务器层,提供事务管理服务 and 规则支持,同时还为系统用户提供访问接口。该服务器层包括工作流机、事务管理和规则支持三个模块。

WIDE 对异常处理进行了充分研究,它使用主动规则技术描述异常情况及其处理方法。WIDE 定义了警告、工作流运行异常和组织异常等三种情况,并分别设计了不同的异常处理策略和函数。最简单的方法就是采用 ECA (Event-Condition-Action) 技术来实现,即当事件发生时,条件满足情况下执行某个动作。WIDE 将系统对异常情况的发生、判断和处理方法写成一条条规则,记录在数据库中。当事件被激活时,系统根据所记录的规则进行处理。

#### (4) 基于状态与活动图的工作流管理系统——Mentor

Mentor 项目的研究是为工作流模型的定义、执行和控制提供一个中间件平台,它采用状态和活动图作为建模规范,使用一个可视化工具 Statemate 作为建模工具,执行活动中所需要的软件可以通过 Statemate

中的扩展接口集成到系统中,用户也可以使用其他过程建模工具来建立模型,Mentor 可将所建的模型自动转化成状态活动图。

活动图描述活动之间的数据流动,状态图规定活动之间控制信息的流动。在状态图中,状态的转换由 ECA 规则驱动。状态图还提供正交组件,允许处于相同层次的两个状态图可互相独立地并行执行。为了支持异构环境下的各种应用程序,系统以 CORBA 产品 Orbix 为底层支持,通过 ORB 传递参数和激活应用程序。系统中所有需要被激活的应用程序都必须包装在 Orbix 中,并提供 IDL 调用接口。

### 3.3 workflow 技术的不足

workflow 技术作为一个新兴的研究领域,还存在着许多问题有待于进一步研究解决。现有的 workflow 系统,主要存在以下不足:

#### (1) workflow 模型方面的不足

目前的工作流管理缺乏一种支持过程定义、过程演讲以及过程分析的形式化模型,尽管出现了不少 workflow 建模方法,但在其模型所具有的描述能力上距离对 workflow 过程本质的描述仍有差距,主要表现在模型元素不丰富,动态性能不理想,仍未离开图形语言的局限。由于缺乏良好的理论基础,workflow 管理系统在应用的许多关键特性上(如过程的柔性定义、过程重用、异常处理等)无法得到保证,问题特别突出表现在现有的 workflow 模型在支持非确定性业务和意外情况的处理上存在着不同程度的局限,即所谓的“僵化”问题,这在很大程度上限制了 workflow 在企业应用中的普及与推广。

#### (2) workflow 管理缺乏统一、可行的标准体系

尽管 WfMC 在 workflow 的概念和接口标准化做出了很多的努力,大多数的接口和标准已经陆续给出。但是,这些标准仅仅是停留在技术层次上,而且重点放在语法上而不是语义上,而在概念层次上,则至今没有统一的意见。

另外,企业在应用 workflow 进行业务流程的运作时,最关心的就是 workflow 系统能否与企业原有的各个应用系统很好的集成,使它们成为一个完整业务流程中的有机组成部分。但是,目前 workflow 管理缺乏一个标准化的集成框架来支持对企业常用的分布式应用的集成,在集成的方式方法上没有统一的标准,很大程度上要受到外部应用的限制。

#### (3) workflow 系统的系统结构和能力上的不足

主要表现在目前的工作流系统不适用于企业级的工作流管理,不同的 workflow 产品之间缺乏互操作性,大多数系统不支持异构系统平台。另外,集中式服务容易

造成瓶颈,使得很难对系统性能进行测量。

除此之外,当前的 workflow 管理系统缺乏健壮性和可用性,对于出错的情况,恢复弹性是极小的,没有备份机制和有效的恢复手段。一个大型的 workflow 管理系统可能包括有几千个用户,几十万个并行的过程,并且分布在一个大型网络的几千个为止。这样,其持续可用性就是非常关键的。现有系统缺乏足够的冗余技术和灵活性来取代出错的部件而不中断系统的功能。

## 4 workflow 产品简介

根据不同 workflow 系统所采用的任务项传递的机制不同,市场上的 workflow 产品可以划分为四类:

1) 基于文件的工作流系统 以共享文件的方式来完成任务项传递。这种类型产品开发得最早、发展最成熟、其产品品种较多。代表产品有 FileNet 的 Visual WorkFlow、IBM 的 FlowMark 和 InConcert。

2) 基于消息的工作流系统 通过用户的电子邮件系统来传递文档信息。这种类型的产品一般都提供一种或多种电子邮件系统的集成接口。代表产品有 Novell 与 FileNet 合作开发的 Ensemble、JetForm 公司的 InTempo、Keyfile 公司的 Keyflow。

3) 基于 Web 的工作流系统 通过 WWW 来实现任务的协作。这一类产品起步较晚(在 1995 年以后),但是发展迅速,其市场前景十分看好。许多供应商纷纷改进原有产品或开发新产品以增加对 Web 的支持。代表产品有 Action Technologies 公司的 Action Works Metro、Ultimus 公司的 Ultimus。

4) 群件与套件系统 虽然这一类产品与上面介绍的三种产品在任务传递方式上有很程度的重叠,但是在这里却有必要把它们单独划分成一类,因为这一类产品都需要依赖于自己系统的应用基础结构,包括消息传递、目录服务、安全管理、数据库与文档管理服务等,它们本身就构成了一个完整的应用开发环境。代表产品有 IBM/Lotus 公司的 Lotus Notes、Microsoft 公司的 Office 和 Exchange、Novell 公司的 Group Wise。

目前,市场上 workflow 产品十分丰富,它们在功能、规模和应用范围上各有特色,在此介绍 FileNet、JetForm、IBM 和 Action 四个公司开发的比较有影响和代表性的 workflow 产品。

### 4.1 FileNet 公司的工作流产品

VisualWorkFlow(以下简称 VW)是 FileNet 公司集成文档管理软件系列产品中的一部分,它与集成文档管理软件中的其他产品合作,可为工作量大、性能要求高的经营过程提供全新的可扩展的过程改进方案。

它允许迅速地评价和改进机构工作方式,可随时查看过程中关键的细节,实时掌握过程的运行情况,还能帮助生成天、周、月、季度或者年的工作负载报告,以便改进过程模型。VW 还提供 Web 服务,支持跨企业经营过程集成。VW 在实时查询过程或者活动的运行状态时,可提供事件日志和统计报告,预计瓶颈可能发生的地点,平衡工作量,优化执行过程,解决工作负担日益增长的问题。它还能够自动确定每个 workflow 活动的参与者在某段时间内应完成的工作,提供有关工作量的信息,并根据活动的执行情况实现相关人员间的任务自动分配。

VW 的解决方案建立在基于组件的软件结构基础之上,它将面向对象技术扩展到 workflow 应用开发中,提高了软件组件的重用率。在其实施中采用先进的工作队列处理方法,保证系统具有较高的可靠性。VW 主要由 VW 服务器、VW 服务、VW 设计模块、VW 执行模块、VW 监控模块组成。它还针对不同的开发环境提供面向不同任务执行程序的开发机制,包括 OLE 接口、ActiveX 控件和一系列 API,用户可以自己选择采用哪种方式与 VW 相连。

在 Workflo 的基础上,FileNET 又发展了基于 J2EE 的 P8BPM 套件。它有一个支持生产 workflow 的过程模型,对生产 workflow 应用有较高的管理能力,还有一个支持终端用户定义和执行特殊 workflow 的网络浏览器。它支持并行工作,具有分布式的结构、高通用性的应用程序接口和较高的企业应用集成能力。

## 4.2 JetForm 公司的工作流产品

InTempo 是 JetForm 公司推出的企业 workflow 产品。它基于 C/S 方式,其客户端相互独立,适于管理型和设定型的经营过程,比如:合同管理、顾客问题解决、销售和预算审批、订购、消费报告、政策回顾、用户调查等等。InTempo 具有以下主要特点:

- 1) 客户端相互独立 企业可针对不同任务选择合适的客户端环境。
- 2) 适用范围广泛 支持各种数据库和消息管理系统。
- 3) 自动分配活动 可自动将任务分配给相应的人员,并使用电子邮件系统传输信息,可使任务自动出现在用户的信箱中,简化了任务接收过程。
- 4) 版本管理 提供了对过程模型的不同版本进行管理的服
- 5) 支持结构化和非结构化的数据 用户可将任何形式的文档附加在过程中传给下一个用户。
- 6) 时间管理 提供了不同程度的时间管理服务。用户可以在创建模型时加入期限限制,并根据用户定

义的时间要求执行超时提醒、过时任务删除等操作。

7) 支持设定型的工作流 允许在 workflow 中加入决策型的活动,对异常活动和人员参与的活动,都将记录在案。可以处理更为复杂的工作流,为企业更快、更准确地决策提供了保证。InTempo 由组织建模、角色建模、过程建模、代理、系统管理、系统监控、Web 连接器等模块组成。

## 4.3 IBM 的工作流产品

IBM MQSeries Workflow 是 IBM 公司推出的 workflow 管理产品,是 IBM 商业集成软件 IBMMQSeries 的一部分。它将经营流程从应用逻辑中分离出来,可以帮助企业用更少的时间、以更快的速度和更高的灵活性集成资源、减少消耗、消除错误、提高生产力,根据市场需求改变经营过程。它以 IBM MQSeries 为基础,将分布在一个异构平台内的不同活动、系统和应用程序有机地集成起来,为在 Internet 环境下实现电子商务提供了保障。MQSeries Workflow 具有以下特点:

- 1) 支持异构系统 可支持 25 种不同的操作系统。
- 2) 自动分配任务 可根据模型定义自动地分配任务。
- 3) 图形化的界面 提供了图形化编辑器,帮助用户快速完成经营过程定义。
- 4) 帮助用户执行和优化过程 能充分利用有关经营过程的知识,帮助用户定义、记录、测试、控制、执行和改进经营过程。

MQSeries Workflow 由 workflow 服务器、管理功能模块、程序执行代理、审查监视功能模块等组成。它还为不同用户提供不同的操作界面。

## 4.4 Action 技术公司的工作流产品

Action 为知识工程师提供了一套基于 Web 的 workflow 管理软件 Action Metro。它不仅能够管理确定的过程,而且对经营过程中不可预见的问题、要求和机遇进行控制,还可为用户提供管理设定型的协作和任务的工具。客户可用 Action 提供的工具,在整个企业甚至企业间建立基于 Web 或 C/S 的 workflow 管理环境。Action 产品将 Internet 技术与 SQL 事务处理技术结合起来,并提供了内置的安全措施,即使用来管理非常复杂的、分布式的和需要不断修改的工作流,也能够保证经营过程的一致性。

与传统的工作流产品相比,Action Metro 更适合于基于知识的工作流。它支持用户与信息之间的交互(如信息获取、用户协商、发现并解决问题等),主要具有以下特点:

1) 实时提供状态信息 Metro 的开放式应用结构, 保证用户及时得到执行活动所需的最新的活动状态、应用数据和应用工具, 并据此作出实时数据和过程状态的决策。

2) 瞬间智能地分配任务 Metro 自动地将任务分类, 用户可通过基于 Web 的工具箱接受任务, Metro 也可以通过 Metro WorkLinks 直接将任务送到用户的电子信箱中。

3) 提供过程一致性服务 Metro 可在多个服务器上同时进行更新, 并在必要的时候进行事务回滚。

4) 功能强大的过程编辑器 Metro 提供图形化的过程编辑器, 用户可方便地定义任何类型的过程模型。

5) 过程评估和改进 Metro 保证对过程的每一步都进行记录、跟踪和说明。能时刻跟踪状态并收集数据, 用户能够发现问题所在, 并对过程模型进行改进, 直到满意为止。

6) 支持 Web 技术 Metro 允许用户通过 Web 参与到 workflow 实例的运行中。Metro 系统包括可扩展的工作管理应用程序、过程建模工具、过程管理工具和一系列接口, 并支持代理执行器为用户自动处理任务。

## 5 workflow 管理系统的发展趋势

近些年, workflow 技术作为一种过程建模和过程管理的核心技术, 在计算机应用领域内获得了迅速发展, 它与其他应用系统结合可以生成符合企业需求的各种业务管理系统, 如文件管理系统、办公自动化系统、项目管理系统、企业资源计划 (ERP) 系统、产品数据管

理 (PDM) 系统、供应链管理 (SCM) 系统。

此外, workflow 技术还被广泛应用到企业业务流程再造 (BPR) 中, BPR 所关心的事对组织机构内部的业务过程进行评价、分析、建模、定义以及操作上的实现。workflow 管理的最大优点是将应用逻辑与过程逻辑分离, 它提供了把业务过程逻辑从具体的 IT 工具的操作中分离出来的方法, 在不修改具体功能的情况下, 通过修改过程模型改变系统功能, 完成对生产经营部分过程或全过程的集成管理, 可有效地把人、信息和应用工具合理地组织在一起, 发挥系统的最大效能。workflow 技术使企业实现对经营管理和生产组织的过程控制以及决策支持, 实现现代企业“在适当的时间把适当的信息传给适当的人”的要求。

workflow 管理系统应该支持异构、自治和分布环境中应用系统的集成和互操作, 提供集成以前的应用系统的方法, 以保护过去的投资, 能够灵活地支持组织机构的改组, 并支持当今有关动态企业 (Dynamic Enterprise) 的技术。具有一定的容错能力, 在有错误产生时 workflow 管理系统能保证 workflow 执行的正确性和可靠性。

随着互联网的普及出现了许多基于 Web 的 workflow 解决方案。通过观察当前商品化的基于 Web 技术的 workflow 管理系统, 显示出大多数产品都仅仅是可以部分地使用 Web, 但是发展的趋势是越来越多地使用面向 Web, 但是由于 Web 及浏览器本身的限制, 只能提供 Client/Server 计算模式, 并且所使用的 CGI 接口只有有限的编程能力, 在位置透明性、支持事务功能、安全性、性能等方面还有待于进一步改善。

## 第4章 产品数据交换技术

### 1 产品数据标准概述

随着计算机技术的发展与成熟,计算机辅助技术在产品设计领域得到了广泛应用。在这个领域里,设计过程的各个阶段所采用的“手段和方法”以及由此产生的“结果”都是以各种各样的数字化“信息”为基础的。为了有效利用这些信息,方便企业内部和企业间的通讯和交流,越来越多的用户需要把产品数据在不同应用系统之间进行交换。因此,有必要建立一个统一的产品信息描述和交换标准,即产品数据交换标准。

#### 1.1 产品数据及产品模型

产品数据是产品对象在整个生命周期中的描述信息。与产品设计阶段(功能确定阶段、原理求解阶段和详图设计阶段)有关的产品数据,按照信息分类可以归纳如下:

1) 描述产品、零部件形状和尺寸的几何/拓扑信息。这里包括二维线框和表面模型,三维线框、表面和实体模型。

2) 含有工程技术说明的图形信息,包括二维和三维图形。

3) 描述零部件基本性能和结构性能的工艺信息,包括几何特性、形状元素、加工性能、装配性能、检验性能。它决定了与制造技术有关的性能。

4) 结构/连接信息,包括几何类、工艺(装配)类,以及几何与图形、几何与制造、几何与设备、几何与检验等的联系。

5) 功能信息,机械中例如运动学、液压、气动等信息。

6) 管理信息,例如状态改变、访问权限、分类、标识等信息。

7) 通信有关信息,例如协议、数据压缩和维护等信息。

8) 特性变化的参数化信息。

9) 用于方法和规则制定的过程/方法信息。

对于机械设计,需要描述的信息有:

1) 产品形状。

2) 加工公差、表面质量和材料特性。

3) 计算结果,例如有限元计算或分析计算的结果。

4) 生产计划,例如生产流程或作业计划的结果。

5) 功能关系和产品结构。

6) 管理与控制数据。

产品模型是对产品数据的组织和描述。模型是对某一事物的一种全面描述。产品模型就是解释有关产品对象在整个产品生命周期中的各种问题的描述。这里所指的生命期不仅包括产品的产生,也就是开发、设计和制造,还包括其他功能如销售、维护以及最终消除或产品的循环利用。因此产品数据描述产品的所有属性,并含有产品制造及其他功能活动所需信息。因此确切地说,“产品模型”是通过数据涉及到的对产品的一种虚拟描述。

#### 1.2 产品数据交换接口

在计算机集成制造中,产品数据的交换充当着一个越来越重要的角色。例如,只要在计算机辅助模型中存在设计数据,则无需经过图形或手工的重复输入或生成,就可以有效地在下游部门(例如直接传入设计分析人员的有限元模型或工艺设计人员的工艺模型)重复使用,并在这些部门生成其他的对产品生产周期有意义的信息(例如有限元分析结果和生产计划所需的工艺规程)。数据的重用简化了输入环节,降低了费用,减少了出错可能,同时,它是实施并行工程和确保不同企业间数据一致性的重要前提。

不同计算机系统间的数据交换,只能通过开发的系统和合适的接口来解决。开发系统允许用户根据功能和经济性提供系统的最佳配置方案,而合适的接口为不同系统间的数据交换提供有效的技术支持。“接口”通常是指在两个系统或系统模块之间,在相互通信过程中进行信息交换的一系列条件、规则和协议。一般来说,应该包括三种接口(如图 39.4-1)。在设计领域,程序接口和描述性接口显然是特别重要的。

不同系统间实现数据交换的接口方式有两种,如图 39.4-2 所示。图 39.4-2a 表示的是专用接口方式,其特点是原理简单,接口处理器易于实现,由此开发的转换效率高。但当系统数增多时,每个系统接口数和接口总数会急剧增加。从昂贵的开发费用来看,为每一种可能的系统结构开发一套特定的接口(处理器)显然是不切实际的。第二种方式是通过标准接口实现数据交换(图 39.4-2b)。从系统的开放性和投资费用的经济性考虑,采用此种方式显然是最明智之举。

接口定义标准化,即不存在解释的二义性,能实现 CAD 系统与其他计算机系统有效的信息交换。单一地定义数据接口具有如下理由:

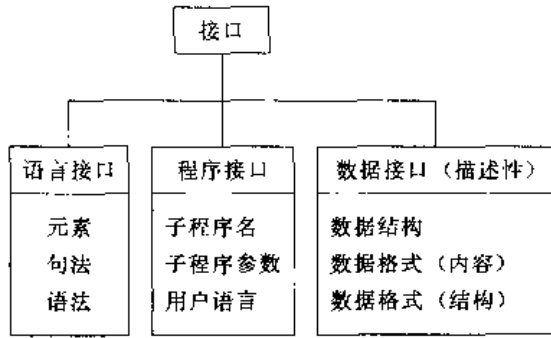


图 39.4-1 接口类型

1) 单一接口可使软件包模块化。因为复杂系统是由许多单个组元连接而成，而这些组元的设计又是相

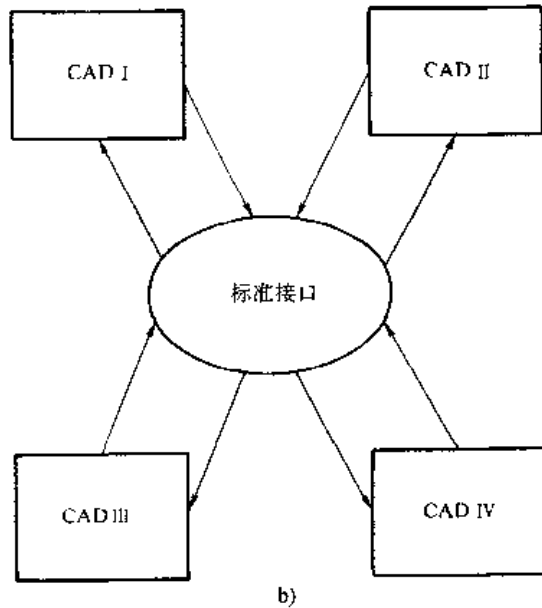
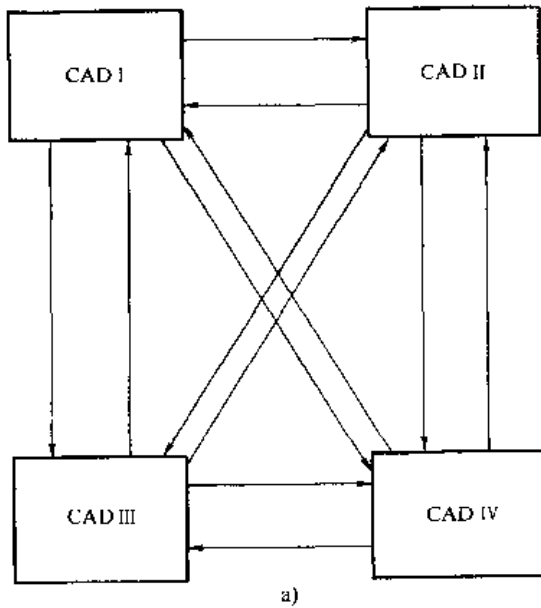


图 39.4-2 有无标准接口的不同系统间信息交换接口的对比

- a) 无标准接口时，每个系统需要  $2(n-1)$  个接口，所有接口数总和为  $n(n-1)$
- b) 有标准接口时，每个系统需要接口数为 2，所有接口数总和为  $2n$

### 1.3 产品数据标准的发展

产品模型数据通过接口进行交换。这里有两个原则具有特殊意义：第一，数据交换应该尽可能地不引起信息的丢失，以确保数据在其他计算机系统中具有进一步的可处理性；第二，确保 CAD 系统的各部分模块化和所属各模块的统一描述。产品数据标准首先是 CAD 系统接口的标准化。

自 20 世纪 50 年代起，美国麻省理工学院积极从事于计算机辅助设计/制造技术的开拓性研究以来，到 70 年代后期，计算机辅助几何造型技术和各种 CAD 系统得到迅速发展，并得到逐步推广应用。随着 CAD 系统应用领域的扩大，CAD 系统数据直接通过计算机进行交换已经在所难免。事实上，具有远见卓识的计算机软件公司、行业及国家标准化组织从 70 年代初就已

互关联的；

2) 通过单一的、用户友好的接口，用户能够编辑和扩展 CAD 系统。因为 CAD 系统常会划分为与专业领域无关的“核”，和与专业领域相关的“壳”，通过接口将构成一个开放系统。

此外，接口定义的标准化，能实现不同 CAD 系统和 CAD 系统组元间的连接，既扩大了企业内的应用范围，并能使数据转换跨越企业界线，同时标准化接口使 CIM 方案能通过标准化软件包的连接来实施。值得指出的是，尽管软、硬件的寿命越来越短，而且为了确保产品的运行可靠，必须经过严格的测试等要求，但是，标准接口能够确保信息处理领域中投资的回收。

经认识到各种 CAD 系统之间需要交换数据的必要性，并积极组织力量或成立专业机构，设立各类研究计划，开展规范的制定、接口与标准的研究与开发工作。

图 39.4-3 描绘了国际上产品数据交换技术发展过程的总貌。图中按时间轴的排列表达了各种不同标准/规范的开发期和发展史；标准/规范间的箭头线描述了各种不同规范/标准之间相互促进、补充与延伸的情况。图中还标出了各种标准/规范的相应组织机构和对标准/规范的发展起主导作用的重大的研究计划项目。下面主要对 IGES 和 SET、VDADIN 等标准做一个介绍。

#### (1) IGES (Initial Graphics Exchange Specification)

IGES，即初始图形交换规范，其初衷是为了能在计算机绘图系统的数据库上进行数据交换。其目的是



制定美国国内各种 CAD/CAM 系统间交换几何信息的统一标准。IGES 从 1979 年开始制定, 1981 年 9 月由美国国家标准协会 (ANSI) 公布为美国标准 IGES1.0 版本, 到 1990 年已公布 5.0 版本。IGES 标准的数据格式和接口结构特别适合于 CAD 系统工程图

面起到过重大作用。但 IGES 占用的数据空间较大, 数据结构也不甚理想, 并且缺少实施和测试的规范, 这使得在数控加工系统中接口的集成产生困难。另外, IGES 图形标准的不完备性还表现在: 元素定义不精确; 元素应用范围不清晰; 度量 and 几何之间没有逻辑关系等。

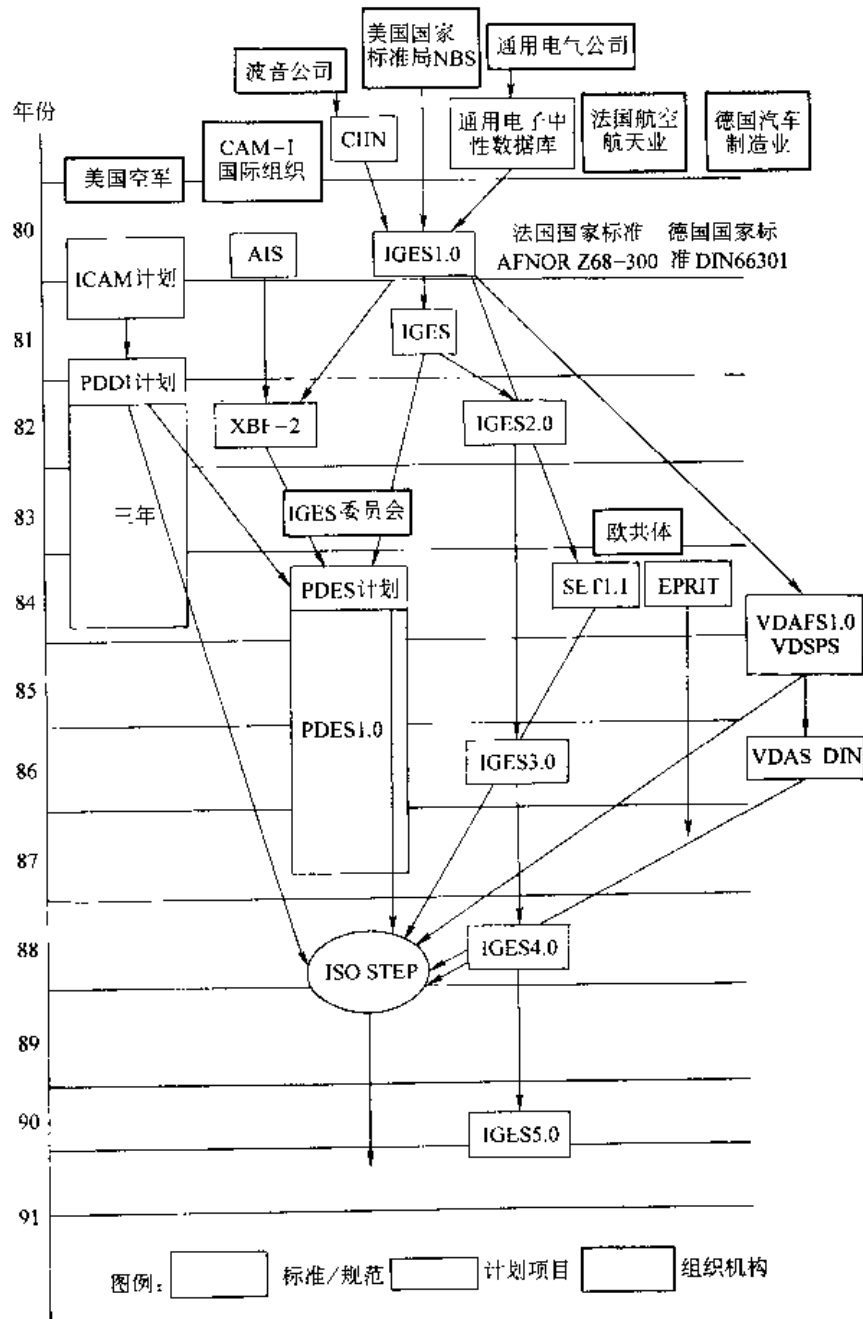


图 39.4-3 国际上产品数据交换技术发展过程的总貌

(2) SET 与 VDADIN

与 IGES 不同版本发展的同时, 由于各个国家不同行业, 在工程应用实践中所暴露出来的数据交换的问题, 发现 IGES 并不完全能满足各自的实际应用需求。这里, 西欧各国在产品数据标准方面针对自己行业

的应用做了大量的工作, 并在自己的工业应用中取得了巨大的成功。这些开发工作主要有法国宇航公司在 IGES 数据模型基础上开发、后来发展成为法国国家标准 AFNORzZ68-300 的产品数据传输标准 SET (Standard d'Exchange et de Transfert); 德国汽车工

业协会开发、后来成为德国国家标准 DIN66301 和 66304 的 VDAFS (Verband der DeutSchen Automobilindustrie-Flachen Schnittstelle) & VDA/PS (VDA/Programm Schnittstelle) 以及 DIN TAP (German Standard for Part Families)。它们的共同特点是根据自己国内的特定行业需求开展工作,并发展成为国家标准。这些标准与 IGES 相比,其功能基本相当。主要区别在于:SET 比 IGES 的存储格式紧凑,允许效率高,灵活而易于扩展;VDA/FS 集中于自由曲面数据的交换;VDA/PS 是基于编程接口标准的开发;DIN TAP 集中于零件族的数据交换。

随着 CAD/CAM/CAE 等应用系统的迅速发展以及应用领域的扩大,带来了信息交换内容及种类的急剧增长。显然,以上这种以几何图形信息交换为初衷的标准/规范,已越来越不能满足各行业、各领域的数据交换的需求。因此,研究新一代产品数据交换的方法、规范/标准势在必行。值得指出的是,IGES、SET、VDAFS&VADPS、ESP 等开发成果和应用实践已经为新一代标准/规范的研究与开发提供了坚实的技术基础和有利条件,在工业应用中不断暴露出的不足以及工业界提出的新需求又为标准/规范的研究提供了推动力。对新一代产品数据交换规范形成和发展起主要影响,并作出巨大贡献的有三个研究计划,它们是 PDDI 计划, PDES 计划和 EXPRIIT 计划。正如图 39.4-3 所示,这些计划后来都成为国际标准化组织 ISO 制定 STEP 的主要内容。

## 2 产品模型数据标准 STEP

正如前面所述,为了传输设计数据,已经有各种接口。但它们都受到单元技术、功能范围和使用可能性的局限。例如目前所使用的接口,如 IGES, VDAIS, VDAFS 或 SET 都未能满足产品模型数据交换接口的要求。它们只适合于以技术图形或简单几何模型的形式在局部的计算机集成制造范围内传输产品信息。然而,更详细的信息,如公差定义、材料属性、产品明细表以至于任务规划等等根本无法传输,或只能通过计算机可解释的格式部分地传输。

国际标准化组织 ISO 已经认识到现有接口的缺陷,从而导致了 STEP 接口的开发。1983 年 12 月,ISO 设立了 184 技术委员会 (TC184),即工业自动化系统。TC184 下设的第四个委员会 (SC4),其工作领域就是产品数据表达与交换。ISO TC184/SC4 制定的标准被称之为产品模型数据交换标准 STEP (Standard for the Exchange of Product Model Data),在 ISO 中代号为 ISO10303。

### 2.1 STEP 标准的特点

STEP 的目标是开发一种国际标准,用来定义产品的模型和传输格式,实现在产品整个生命周期中计算机可解释的产品信息的无二义性表达。属于产品生命周期的信息来自设计部门、生产部门、产品检验以及维修和维护部门。这些部门还提出了对所传输的局部模型的要求,例如这些模型的拓扑/几何对象描述、公差、材料特性和表面性能、组织管理数据、计算和分析数据、工艺计划数据、产品结构数据和描述数据等等。

STEP 的开发要求还包括数据的完整性、存档能力、可扩展性、有效性、与其他标准的兼容性、最少的实体类型、与运行系统的不相关性、与数据内容的独立性、物理存储格式,以及构造专用的应用数据子集的能力(也称之为子集的构成能力)等。同时,随着工程实际中所开发接口的标准化,为了消除现有标准所暴露的缺陷,如不充分的形式化描述、非指令性的测试准则等等,STEP 已发展成为一门方法学。

STEP 开发的基本原则是从组织和逻辑上分为三个层次。

第一层为应用层,描述了各种应用的参考模型。它采用的工具是描述信息单元(即实体)及其关系的 IDEF 图形表示法。

第二层为逻辑层,产生的结果是 STEP 实体、STEP 属性和对这些属性有效条件的一种格式化描述图。按 STEP 标准方案,在第一层集成各种应用,然后将集成模型传至第二层。在第二层,实体的描述采用专门为之开发的形式化描述语言 EXPRESS。

第三层为物理层,它描述的是一种物理文件格式。

### 2.2 STEP 的体系结构

图 39.4-4 是 STEP 的体系结构,它反映了用于 STEP 开发的原理和方法。从图中可以看出,STEP 表现为一系列的 ISO10303 标准,除了产品模型外,还定义了描述方法(见图中左侧部分)、实现方法(见图中中间部分底部)以及一个一致性测试方法学和框架(见图中右侧部分)。STEP 的资源模型可视为一系列的可构造块(见图中中间部分),利用这些构造块可以按照标准化规则和方法来定义特定应用领域的产品模型(应用协议)。

对应前面已经提及的 STEP 开发基本原则的三个层次,应用层包括应用协议及对应的抽象测试集,它是面向具体的不同领域的应用,是与应用有关的一个层次。第二层的逻辑层包括集成资源,是一个完整的产品数据模型,它是从实际应用中抽象出来并与具体实现无关。最底层是物理层,包括实现方法,给出具体在计算机上的实现形式。

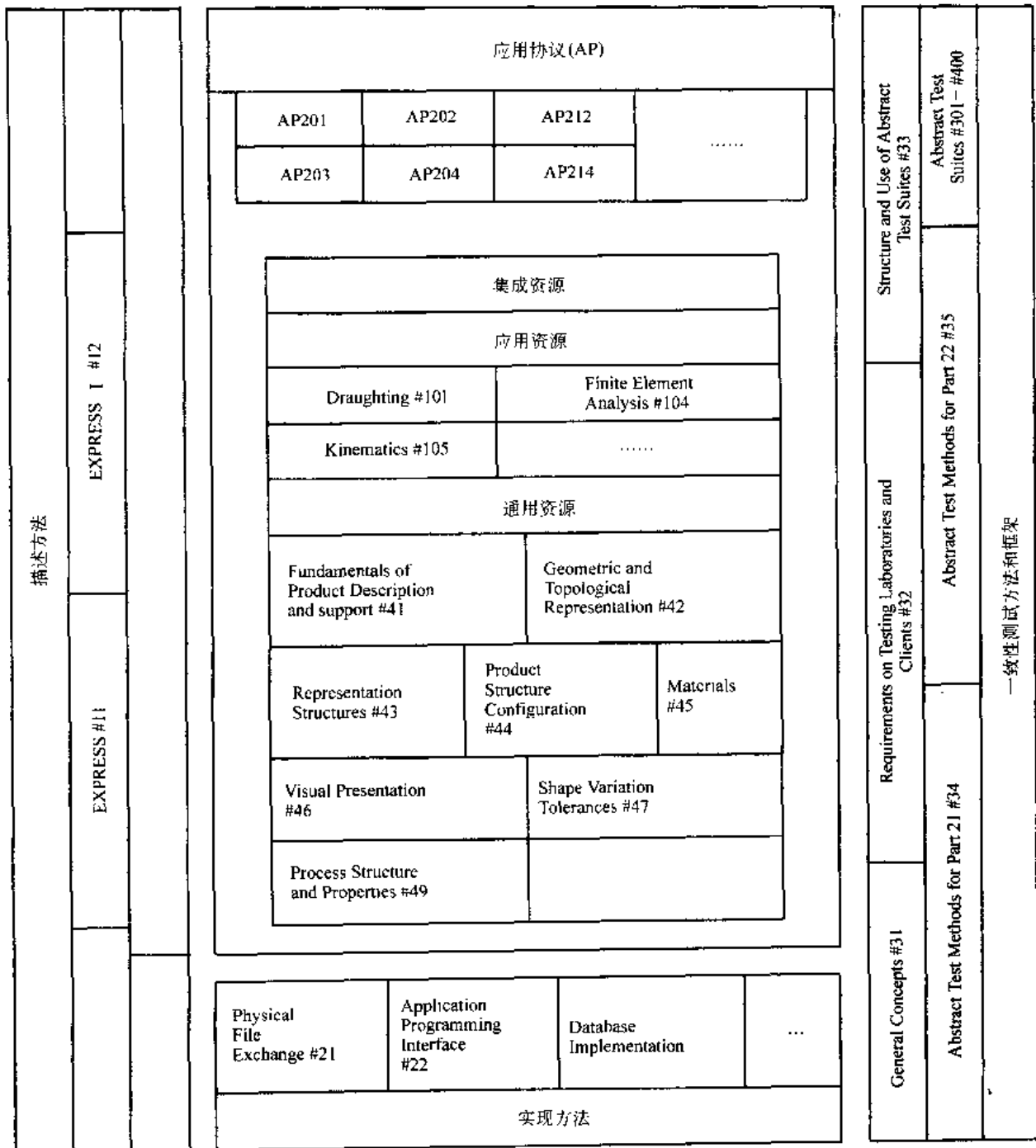


图 39.4-4 STEP 的体系结构

(1) 描述方法

描述方法是 STEP 标准的基础，由模型描述语言和通用定义规范组成，是建立 STEP 产品数据模型的工具。标准编号由 1 到 19，如：

ISO10303-1 综述与基本原理 (ISO)；

ISO10303-11 EXPRESS 语言参考手册 (ISO)。

EXPRESS 语言中规定了数据元素、约束、关系、规则和函数等，用户可以根据需求利用它们定义新的

资源构件，也可以通过它们对已有的资源构件（如集成信息资源中的资源构件）进行修改（称之为解释）。EXPRESS 语言规定以模式 (SCHEMA) 作为描述信息模型的最外层框架，模式内包括模式间引用关系以及实体、类型、常量、函数、过程、规则等说明，通过这些说明定义事物间的联系，从而建立信息模型。因此，EXPRESS 语言是一种具有丰富表达能力的信息建模语言。

## (2) 集成信息资源

集成信息资源位于 STEP 逻辑层, 包括 STEP 推荐的概念模型, 是 STEP 的核心部分。其中的资源构件是描述产品信息模型的基础。其使用贯穿于应用协议组 (APs) 的各标准, 并且促进应用协议的集成性和相互协调性。集成信息资源可细分为集成通用资源、集成应用资源和应用解释构件。其中集成通用资源被用于整个 STEP 应用协议的领域中, 标准编号由 40 到 49。集成应用资源的作用域相对狭窄, 只能被应用它的应用协议使用, 标准标号由 100 到 199。当两个或两个以上 AP 的信息需求出现重复时, 必须使用相同的解释, 即在 AIM 中使用能够满足这些信息需求的相同的集成资源构件和定义。在这种情况下, 引入了应用解释构件 (Application Interpreted Constructs) 的概念。应用解释构件主要为建立应用协议提供资源引用和解释的标准, 标准标号由 500 开始, 如:

### 1) 集成通用资源

ISO10303-41 产品描述和支持基本原理 (ISO);  
 ISO10303-42 几何与拓扑表达 (ISO);  
 ISO10303-43 表达结构 (ISO);  
 ISO10303-44 产品结构配置 (ISO);  
 ISO10303-45 材料 (DIS);  
 ISO10303-46 视觉展现 (ISO);  
 ISO10303-47 形状公差 (DIS);  
 ISO10303-48 形状特征 (撤销);  
 ISO10303-49 过程结构和特性 (DIS)。

### 2) 集成应用资源

ISO10303-101 绘图 (ISO);  
 ISO10303-102 船舶结构 (WD);  
 ISO10303-103 电子电子连接 (撤销);  
 ISO10303-104 有限元分析 (CD);  
 ISO10303-105 运动学 (ISO);  
 ISO10303-106 建筑构造核心模型 (WD)。

### 3) 应用解释构件

ISO10303-501 基于边的线框 (CD);  
 ISO10303-502 基于壳的线框 (CD);  
 ISO10303-503 几何边界二维线框 (CD);  
 ISO10303-504 绘图标注 (CD);  
 ISO10303-505 绘图结构和尺寸 (CD);  
 ISO10303-506 绘图元素 (CD);  
 ISO10303-507 几何边界曲面 (CD);  
 ISO10303-513 基于边界表达 (CD);  
 ISO10303-514 高级边界表达 (CD);  
 ISO10303-515 构造实体几何 (CD)。

## (3) 应用协议

应用协议位于 STEP 的应用层, 用于描述特定产

品实际应用的数据模型。它们不仅阐述了用什么数据描述一个产品, 而且描述了该数据是如何被用在模型中的。标准编号为 200 到 299, 如:

ISO10303-201 显式绘图 (ISO);  
 ISO10303-202 相关绘图 (ISO);  
 ISO10303-203 配置控制设计 (ISO);  
 ISO10303-204 用边界表达的机械设计 (CD);  
 ISO10303-205 用曲面表达的机械设计 (CD);  
 ISO10303-213 加工件的数控工艺规划 (DIS);  
 ISO10303-214 汽车设计过程的核心数据 (CD)。

## (4) 实现方法

实现方法描述了 STEP 标准的不同实现方式, 如:  
 ISO10303-21 交换文件结构的纯正文编码 (ISO);  
 ISO10303-22 标准数据访问接口 (DIS);  
 ISO10303-23SDAI C++ 语言联编 (CD);  
 ISO10303-24SDAI C 语言联编 (CD);  
 ISO10303-26SDAI 接口定义语言 (IDL) 联编 (WD)。

## (5) 一致性测试

一致性测试提供了关于产品对 STEP 标准的一致性测试方法的信息, 一般性测试方法学和框架, 抽象测试集, 为开发抽象测试套件提供指南并描述了测试实验, 如:

ISO10303-31 一致性测试方法和框架 (ISO);  
 ISO10303-32 测试实验室和客户需求 (DIS);  
 ISO10303-33 抽象测试套件的结构和使用 (CD);  
 ISO10303-34 抽象测试方法 (CD);  
 ISO10303-35SDAI 实现的抽象测试方法 (WD)。

## 2.3 EXPRESS 描述语言

集成资源和应用协议中的产品数据描述要求使用形式化的数据规范语言来保证描述的一致性、无冲突和语义上的无二义性。这种形式化语言既具有可读性, 使人们能够理解其中的含义, 又具有能够被计算机理解的形式化程度, 有利于计算机应用程序和支撑软件的生成。因此在 STEP 描述方法中定义了 EXPRESS 规范描述语言和它的图形化表达 EXPRESS-G。

EXPRESS 是一种形式化信息建模语言而不是一种编程语言。它是 STEP 系列中的一个部分内容, 用以描述 STEP 中其他部分的信息要求, 且标准化为 ISO10303-11。其设计目标包括:

- 1) 语言不仅能够为人所理解, 而且便于计算机处理。
- 2) 语言能够区分 STEP 涉及纷呈复杂的内容。
- 3) 语言的重点放在实体 (Entity) 定义上。实体定义包括实体属性和这些属性上的约束条件。

4) 语言尽量与具体实现无关。

EXPRESS 语言吸收了许多语言的功能和特点,并增加了一些新的功能,以便更适宜于表达信息模型。需要指出的是,EXPRESS 语言不是一种程序设计语言,它不包含输入/输出、信息处理、异常处理等语言元素。它是一种具有面向对象特性的描述语言。EXPRESS 通过数据和约束,清楚、简明地定义了对象,给出了一个概念模式。因此 EXPRESS 不仅是产品数据模型的规范语言,用来描述集成资源和应用协议,而且是整个 STEP 中数据模型的形式化描述工具。这种用形式化语言描述标准,为标准在计算机的实现提供了良好的基础,可视为计算机相关国际标准制定工作中的一种新突破。EXPRESS 同样也适用于 STEP 范围以外的信息建模。

## 2.4 STEP 标准化状况

已注册成为国际标准的分册内容如表 39.4-1 所示。这些内容,特别是应用协议,将随着国际标准化进程不断地延伸。

表 39.4-1 已注册成为国际标准的分册

分册	说 明
1	概况及基本原理
11	EXPRESS 语言参考手册
21	交换结构的文本编码
31	通用概念(一致性测试方法学)
41	产品描述基础和支持
42	几何和拓扑表达
43	表达结构
44	产品结构配置
46	视觉展现
101	绘图
105	运动学
201	显式绘图
202	相关绘图
203	配置控制设计

## 3 主要应用协议介绍

### 3.1 电器设计和装配应用协议 AP212

应用协议 212 电气设计和装配是 STEP 技术在电气和汽车工业所迈出的第一步。它由 proSTEP 和 Siemens 公司与 DKE 和其他国际合作伙伴共同开发,并在 ISO TC184/SC4 和 IEC TC3 综合框架下,形成了 ISO/IEC 标准 10303 212。

AP212 就是给出电器工业产品开发过程链的产品数据的一种描述。因此该模型是实现电器产品进行产品数据和信息交换有效、经济的前提。这就是说,产品数据的交换不仅在制造商和它们的客户、供应商之间进行,而且还可以在公司内部进行。与电气系统和设备(例如发电厂,电力输送或分布电网,汽车,船舶,建筑)相联系,协议中包括了工程图,模式表达(如功能规划,电路图,装配图),网络表,连接表和元器件列表所必需的信息。显然,这些信息的可获取性和共享对电气系统的构造,维护,服务或检测具有重大的意义。

#### (1) AP212 辖域

就目前工业界中所使用的接口格式而言,如 IGES, EDIF, SET, VNS 或 DXF,通常只允许以工程图的形式交换几何信息。AP212 则考虑了电器产品的电或某些机械特性。因此,除几何描述外,它还允许电路逻辑的完备描述,以及后续的工艺和非工艺数据描述。

为能满足多种需求,AP212 支持下列的产品数据类型:

- 产品及其结构的通用描述;
- 功能及其结构描述;
- 连接及网络;
- 信息;
- 安装、布线和配线;
- 项标识和归类,如对资源、端口和信息;
- 属性、特性和注释;
- 文档和图形表达;
- 版本和配置管理;
- 审批、工作单。

由于每个对象有各自不同的特性,AP212 覆盖了如面向物理产品的、面向功能的或面向位置的产品特性。对于完备地描述一个电气系统,其他重要要素将在文档、2D 表达模式和产品数据管理的模型部分中加以考虑。

#### (2) 验证

应用协议开发的一个重要任务是模型的验证。在该过程中,模型中的内容将在不同的企业用户和系统开发商的合作下,就其完备性、精确性、综合性等方面进行详细的检测。在验证过程出现的与模型结果相违背的问题记录在案形成文档并归入 AP 中。验证的基本过程如图 39.4-5 所示。

在 Pro STEP 的领导下,AP212 的未来用户——来自汽车、化工、电子和航空工业以及各应用部门和系统开发商共同参与了验证工作,讨论了在不同企业和项目中的需求。所有来自于验证过程中各种需求都将

补充到新的 AP212 文件中去。为了支持 AP212 标准化的进一步发展,验证工作将继续进行。

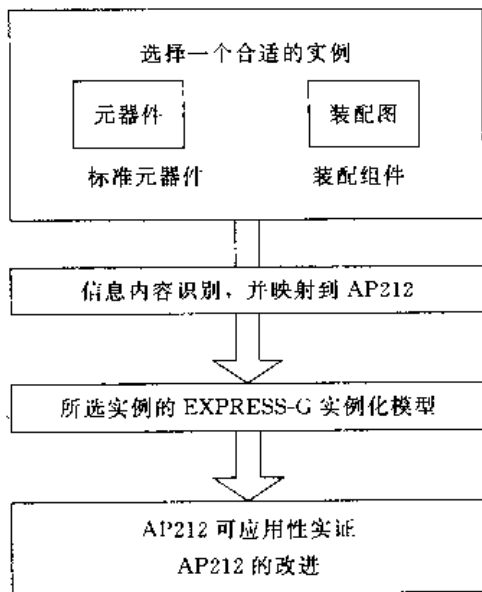


图 39.4-5 AP212 的验证过程

### 3.2 汽车机械设计过程 AP214

在德国汽车协会 (VDA), pro STEP 和其他国际合作伙伴合作下,汽车工程设计过程的应用协议核心数据形成 ISO 标准 10303-214。

AP214 描述了在汽车工业中,开发过程链的产品和生产设施数据。

在标准系列 ISO10303 开发的应用协议中,AP214 描述了与机械开发过程相关的技术信息的数据结构。其目的是为确保产品数据在开发过程链中能交换、共享、解释以及进一步的处理。目前这仅涉及到汽车工业中技术数据处理的约 100 家 CA 系统。

除了 CAD 特定数据,如几何表达、图形显示以及工程绘图以外,下列范围的信息可以通过基于 AP214 进行数据结构的表达:

基本的产品定义(如发布和变更的管理信息)和产品结构信息(如装配、零件表, BOM):

- 产品说明、配置管理和工作单;
- 工艺规程;
- 形状特征;
- 技术特性,如质量、惯性矩或材料以及回收特性;
- 公差和表面特性;
- 仿真数据(如运动学)。

在有关 AP214 的开发工作方面,已经形成了一个参考数据模型草案,并呈交给参与伙伴及德国国家标准部门审议。同时,应用参考模型由德国、法国、日本、瑞典和美国的工业实际的案例进行详细的检测。

AP214 于 1997 年 10 月已成为委员会草案 (CD)。

### 3.3 工艺设计的机加工产品定义应用协议 AP224

STEP 应用协议 Part224, 全称“基于加工特征的工艺设计机加工产品定义”,定义了加工单个机械零件的信息表达和交换,信息包括零件特性、过程控制、加工描述、管理数据和需求等。

零件形状的特性是协议 Part224 的核心内容。协议确定了制造中产品形状的独特特点,用以定义制造特征。形状既可由 B-rep 实体模型描述,也可由加工特征描述。定义制造特征的目的是为了便于制造形状的认定,使之可被人和计算机理解。制造特征允许其形状特征用于工艺计划系统决策。产品材料信息决定了工装及材料需求。管理信息用于跟踪有关零件的用户和供应商信息,同时控制制造活动以支持工艺计划。特定管理信息的跟踪是建立工艺计划中反馈控制机制的一个组成部分。

#### (1) AP224 辖域

AP224 指明了如何使用 STEP 集成资源来满足加工单个机械零件的信息表达和交换需求。产品数据基于现有的由加工特征表达的零件设计。下列范围的信息可以通过基于 AP224 进行数据结构的表达:

- 通过机加工工艺制造的单个机械零件;
- 通过铣或车过程制造的产品;
- 为定义制造所需形状的加工特征;
- 构建形状特征所需的特征定义项;
- 用户订购管理数据——跟踪零件从定单到车间的过程,但不包括车间层指令;
- 零件工作指令状态的认定;
- 跟踪原材料状态,记录零件的制造历史;
- 跟踪一个零件的设计异常。

AP224 不支持对下列信息的数据描述:

- 来自工艺计划的结果;
- 装配描述;
- 复合材料零件;
- 板材零件;
- 零件簇;
- 零件设计特征;
- 在制造过程中实现工作指令的计划表;
- 配置。

#### (2) 信息构成

工艺计划信息需求由一组功能单元 UoFs、应用对象和应用声明构成。其中,应用声明描述了应用对象之间的关系、关系的重要性以及保证应用对象的一致性和有效性的规则。Part224 包含了以下 10 个功能单

元。

1) 设计异常: 包含工艺计划生成过程中出现任何异常的错误报告以及异常相应的解决方法, 作为再工艺计划输入的重新生成信息。

2) 特征定义项: 包含构造加工特征所需的信息。明确了加工特征与形状要素 (shape aspect) 之间的关系。

3) 特征截面, 包含识别 2D 形状所需的信息。3D 特征通过特征截面沿一定的路径扫描得到。

4) 制造特征: 用于识别零件形状。代表从零件加工中被去除的具有, 一定体积的材料。

5) 制造零件属性: 包含被定义零件的特性的描述, 这些描述指明了制造需求。制造需求不仅将应用于零件加工前或加工后特定时间的零件状态, 而且将应用于零件制造过程。

6) 制造过程控制文件: 包含零件及其产品定货数据信息以及来自客户本身或客户以外及制造内部跟踪零件制造过程的操作规程的工作指令信息。

7) 制造过程需求文件: 提供与产品数据直接有关的描述文本, 支持产品定义。这些文本明确了作用在被制造零件上的操作或制造过程中特定阶段的零件状态属性。文件可以 (但非必要) 是公司、国防、国家或国际标准。

8) 测试控制: 描述了一个零件形状要素之间或一个零件形状要素与一个不包含该零件形状的引用形状之间测试关系的重要尺寸, 以及制造允许的偏差。

9) 零件管理数据: 包含了管理产品数据所需的信息描述。

10) 零件模型: 包含用来确定作为工艺计划功能模块输入的零件和确定该零件属性联合所需的信息。

### (3) 目前状态

AP224 于 1993 年 10 月 1 日作为工作草案 (WD) 推出, 1995 年 6 月 29 日确认为国际委员会草案 (CD), 现在已作为国际标准草案 (DIS)。

## 4 STEP 工业应用

ISO 标准化委员会在 STEP 的发展策略和标准制订方面已经投入巨大人力, 做了大量工作, 但是仅仅拥有策略和标准是远远不够的。伴随着高技术含量的应用软件、各种开发工具的不断涌现, 在工业界中进行“STEP 技术产品”的快速推广就成为必不可少的重要一环。从 IGES 的应用中, 人们得到的一个教训是, 只有通过用户和软件开发商之间的强力合作才能成功地推广一个标准。因此, STEP 广泛的工业应用, 也必须通过不同开发商的密切合作, 借助相兼容的 STEP 实现给以保证。另一方面, STEP 的工业用户也需要了解

为支持其应用, 哪个 AP (应用协议) 是必须的。但是它并不想处理太多的 STEP 细节。它们使用 STEP 是为了改善其内部的业务过程。

以汽车工业为例, 工业界各企业广泛地应用着种类繁多的 CAD/CAM 系统。这种在一个公司内部, 以及在一个汽车行业范围多种 CAD/CAM 系统共存既是 CAD 技术发展的必然, 也是企业发展过程中某种特定需求所造成的客观现实。为了完成一部整车的开发, 主机厂与其供应商存在着广泛的协作, 形成了一个复杂的供应链。在这个供应链中, 存在着产品数据的频繁变换, STEP 将为这些应用提供很好的解决方案。

### 4.1 Pro STEP 中心

德国 Pro STEP 中心在 STEP 工业应用方面处于世界各国的领先地位。

为满足产品数据交换的需求, 在汽车及电子/电气工业界中具有远见卓识的带头企业发起成立了 Pro STEP 中心的倡议, 并促成了它的建立。Pro STEP 中心的目的是支持与 CA 开发商在开发兼容产品方面的合作, 帮助 CA 用户推广 STEP 以及交流它们在标准过程中的需求。同时, Pro STEP 中心增进了最终用户和系统开发商之间的紧密合作, 提高它们产品的兼容性。

Pro STEP 中心包含两个法定部门——Pro STEP 协会和 Pro STEP 公司。

#### (1) Pro STEP 协会

Pro STEP 协会成立于 1993 年, 协会最初的 38 位成员迅速发展成目前来自 45 个国家的 187 位成员, 其中包括许多跨国公司 (至 1998 年 1 月)。协会成员主要来自代表着科学技术发展迅速的先进工业部门 (如汽车工业和航空工业), 主机制造商和广泛的供应商形成了一个集成的商务社团, 并使协会中的系统开发商成为强劲的伙伴。

在 Pro STEP 协会中, 他们共同追求的目标是优化产品在开发和制造过程中的信息处理过程。从而降低企业内部以及企业间产品数据的频繁交换和复杂管理的成本。为了实现这个目标, 协会不仅提倡在国际标准基础上进行产品数据的描述, 而且也十分注重这些标准在工业应用中的实践。

为 STEP 建立用户驱动的质量保证是协会的另一目标。Pro STEP 协会的一个重要活动是对 STEP 处理器开发商举行圆桌会议进行协商。圆桌会议将系统开发商和用户集合起来, 在软件开发商的共同参与下, 确保基于标准并相一致的 STEP 处理器的开发。

#### (2) Pro STEP 公司

Pro STEP 公司是实践上述协会运作的助手。它将提供 STEP 技术产品和服务, 以此将 STEP 推广到

企业。这些产品和服务包括:

研究和开发各种项目;

项目咨询;

标准软件模块的开发;

面向最终用户的 CAD 数据和产品管理数据交换的工具;

数据交换支持;

培训、教育和服务。

同时,还包括当今数据交换标准如 IGES 和 VDAFS 的支持产品,基于 STEP 的软件开发案例以及专用软件开发等等。

## 4.2 AP212 和 AP214 的工业推广

### 4.2.1 AP212 工业推广

为了将 STEP 数据模型推广到实际应用,针对应用协议 AP212 开展了许多 AP212 的验证和推广项目。

#### (1) EI 项目

EI 项目于 1997 年初由 DASA, debis, Delphi, Ford, Mentor Graphics, Harness Software Ltd, Pro STEP 和 Siemens 等国际性的 E-CAD 系统用户或开发商发起。其主要精力集中在基于 STEP-AP212 处理器的开发以及在开发过程中对这些处理器的测试。该项目任务主要有:

按用户最迫切的需求定义处理器的功能;

开发基于 AP212 的处理器;

应用 AP212 处理器进行产品数据交换的演示;

在工业实践中推广 AP212 技术;

进一步实施 AP212 的建议。

#### (2) "Process chain car-electric/AP212" 工作组

在德国,由德国汽车协会(VDA)和 Pro STEP 公司成立了"Process chain car-electric/AP212",联合工作组,其成员有 BMW, Daimler-Benz, Delphi, Draexlmaier, John Deere, Opel, Siemens 以及 VW。该工作组根据汽车制造商及其供应商的应用需求进行 AP212 的验证

(3) AP212 在能源行业的产品文档领域中应用的可能性分析

为了评估 AP212 在能源分配业务中有关配电系统感应和控制技术产品数据的可用性,有能源分配专门小组(IG-EVU)和 WeidIEluder 公司共同发起进行了调研分析。结果表明,所选范围内的电路图手册几乎全能映射。因此可以说,AP212 能用于动力配电系统感应和控制技术的描述和文档。当然,其前提所开发的 AP212 处理器必须有效。

另外,在对德国 VNS 标准(DIN V40950)的确认

中表明,AP212 包含了 VNS 的功能,因而也就有可能在未来将由 AP212 代替 VNS 标准。

(4) STEP 在化学工业的仪表及控制系统中的应用

另一个 AP212 使用的分析项目是关于化学工业中仪表和控制系统的产品数据交换。为了确认 AP212 是否适合电气系统的产品数据交换,在该项目中对 AP212 的产品模型和常用的 CAE 系统进行了比较。分析表明,AP212 可以表达这些产品信息。因而也就用于 CAE 系统间典型电气产品数据交换。

### 4.2.2 AP214 工业推广

AP214 和 AP203 转换器已经在许多著名的大公司(如 BMW, Bosch, Delphi, General Motors, Hazet, John Deere, Kostal, Mercedes Benz, Opel, Siemens 和 VDO 等)与供应商和客户间的数据交换中获得了工业应用,同时,在公司内部各部门也展开数据交换的应用。在飞机工业领域中,STEP 转换器已用于 BMW, Rolls Royce, Boeing 和 MC Donnell Douglas 等公司。

AP214 重要项目的实例有以下几个。

#### (1) Pro STEP AP214 圆桌委员会

Pro STEP 圆桌委员会是 pro STEP 协会由许多 CAD 系统开发商共同参与的永久性机构。其目的为:STEP 转换器(特别是 AP214)实施上达成一致;STEP 转换器开发的协调以及最佳协作方式的采用;

为能实际应用,给 AP214 标准化以影响和建议。

#### (2) ISAP 项目

ISAP 项目(国际 STEP 汽车项目)起始于 1995 年年中。它分为 ISAP1, ISAP2, ISAP3 三个层次。三家伙伴有 BMW, vw, Opel, Daimler Benz, GM, Ford, Siemens, Bosch, Delphi, Bertrand, ITT Automotive, Delco 和 Molex。Pro STEP 公司协调欧洲市场,而 PDES 公司支持在美国的各项活动。该项目的目的是 STEP 转换器和其他工业应用系统兼容性的质量保证问题。为了实现这个目标,在主机制造商和其供应商之间进行真实数据的交换,并将在不同应用协议(如 AP203, AP214)中所作的测试结果在项目成员中公布、陈述及讨论。

#### (3) STEP-一模型公司

STEP-一模型公司项目起始于 1995 年。其目的是将 AP214 推广到 KUKA 机床制造公司的实际应用中。该项目的工作主要集中在公司内部及外部(供应商和主机厂)的几何和管理产品数据的交换。此外,测试市场上已有商品化的 PDM 系统对 AP214 数据模型的适应性。



## 第 5 章 并行设计的使能工具

### 1 质量功能配置 (QFD, Quality Function Deployment)

质量功能配置是一种系统化的产品质量规划方法。它采用质量屋 (House of Quality, 简称 HOQ) 的形式, 通过定义“做什么”和“如何做”将顾客需求 (Voice of Customer, 简称 VOC) 逐步展开, 分层转换为产品工程特性、零件特性、工艺计划和生产控制等, 形成从用户需求到产品上市的连续转换和“实施流”以确保用户需求就是开发目标, 加强了产品开发各阶段中项目组成员对用户需求及为满足用户需求所采取的措施的理解, 从而提高了产品设计质量。它的特点是将注意力集中于规划和问题的预防上, 而不是仅仅集中于问题的解决上, 是实现产品质量保证的重要手段和工具。

图 39.5-1 表示了质量屋的一般形式, 它清晰地表达了各要素之间的相互关系。质量屋由五个部分组成: A 区是用户需求及权重矩阵, 收集各种用户需求要素, 并比较其重要程度 (相对权重的总和为 1), 逐一填入该矩阵; B 区是工程特性矩阵, 分析用户需求, 确定影响用户需求的功能特性要素, 填入该矩阵; C 区是用户需求与工程特性关系矩阵, 逐一审核各功能要素与用户需求的关联程度, 确定出在单一需求下各功能要素的相对重要性, 填入对应的方格中; F 区是工程特性相关矩阵, 用与检查工程特性各项之间的关联和冲突, 如果某一项对应参数的增加将导致另一项对应

参数的增加, 则在工程特性相关矩阵中填入正相关符号 (+), 反之填入负相关符号 (-), 据此可检查出各项之间是否有冲突, 以及是否可以进行优化; D 区是评估结果矩阵, 表示功能特征项的重要性程度, 根据需求要素的权重和功能特征要素对其影响程度的比率确定; E 区是市场竞争力评估矩阵, 收集主要竞争对手有关产品的信息, 分别评估客户对各种需求的满意程度。该项表示了竞争对手满足各项用户需求的实力, 以确定市场营销的策略。

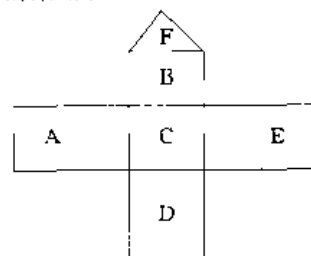


图 39.5-1 质量屋的一般形式

质量功能配置通常分为四个基本阶段, 即产品规划、结构和零件设计、工艺规划、作业控制等, 四个阶段的质量屋的结构形式类似。但产品规划质量屋中的工程特性及其权重可以作为结构设计阶段的需求输入, 由此确定第二阶段为满足这些需求而采取的措施, 构造出第二个质量屋。这样逐个将有关需求映射到产品生命周期各个阶段的有关技术和措施, 形成一个瀑布式逐级分解过程, 并能逐级回溯和审核。图 39.5-2 表示了产品生命周期 QFD 系列及产品需求间的映射过程。

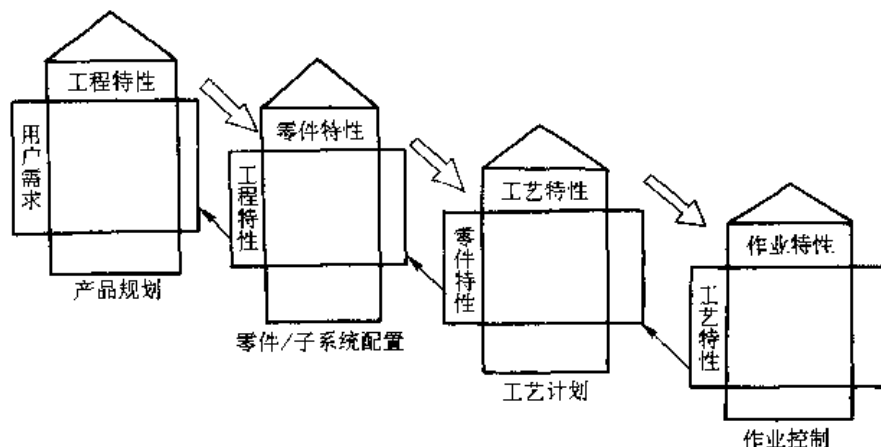


图 39.5-2 产品生命周期 QFD 系列及产品需求间的映射过程

这儿介绍的仅仅是质量功能配置最基本的原理及实施方法。实际应用中, 由于许多信息的模糊性, 要使

信息处理科学化, 还有待大量的研究工作。目前国内外学者的研究重点包括 QFD 模型, QFD 中定性信息的

模糊处理, QFD 自动化实现技术。

## 2 面向制造的设计 (DFM, Design for Manufacturing)

DFM 是面向并行工程的一种设计方法, 主要思想是在产品的早期设计阶段考虑制造因素的约束(例如, 产品制造所需要的机床设备、工装模具、加工工具、测量工具以及相应的时间、费用等), 并及时提供给设计人员, 作为设计、修改方案的基本依据, 在满足功能要求的前提下, 使产品能够快速地、经济地制造, 提高产品设计、制造的一次成功率, 达到降低产品成本, 缩短产品开发周期的目的。

传统的产品设计方法如图 39.5-3a, 产品设计和工艺设计作为两个独立的过程处理, 两个设计过程是顺序执行的。DFM 打破传统的顺序设计方法, 将产品设计和工艺设计并行开发, 如图 39.5-3b, 充分考虑各种制造约束条件、加工条件、装夹方案、工装设计和零件标准化等, 对设计的方案进行评价和决策, 并且及时将评价结果反馈到设计者, 指导产品的设计和提出修改建议。

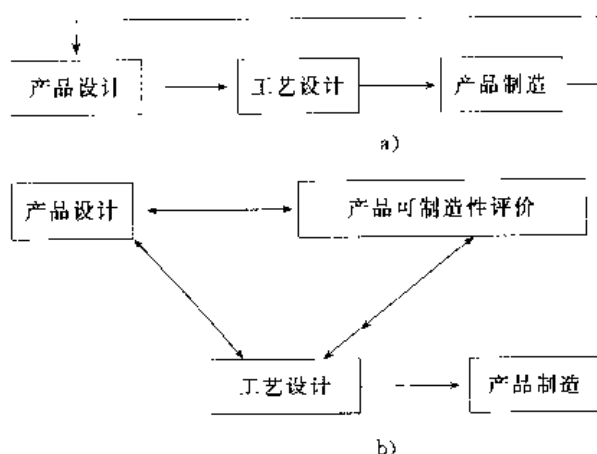


图 39.5-3 DFM 设计方法与传统设计方法的对比

a) 传统设计方法 b) DFM 产品设计方法

从设计过程和设计内容分, 有狭义的 DFM 和广义的 DFM。前者一般被称为面向加工的设计 DFF (Design for Fabrication), 其重点在于通过对零件的结构工艺性、加工经济性和对制造资源的要求的分析评估, 改善零件的结构, 使其在现有的制造环境下易于加工。后者包括产品结构和零件结构的设计, 即 DFF 和 DFA, DFA (Design for Assembly) 的重点在于通过合并零件, 减少零件数量来简化产品结构, 提高可装配性, 降低装配成本。DFM 的实质是在产品设计的各个阶段, 在保证产品功能的前提下, 要充分考虑到设计的可制造性和可装配性等, 并使产品开发的时间、成本和质量收敛于期望值内。

零件级 DFM 主要是根据零件制造(如工艺规划、工装、加工等)的要求对设计进行评价, 并指导再设计。

1) 与资源无关的结构工艺性问题 DFM 系统主要根据结构工艺知识对零件的几何和精度信息进行分析, 对待加工特征的加工可能性、加工容易程度进行评估。

2) 与资源有关的可制造性问题 它是从企业的制造资源能力来考虑的。例如零件的精度和尺寸是否超过现有机床的加工能力, 零件上的孔是否有相应直径的孔加工刀具等等。

3) 零件的制造经济性问题 它要求 DFM 在设计阶段就对零件及其加工成本等进行粗略的相对的估算, 用以指导设计者有针对性地修改设计, 以及对不同设计方案进行评价和取舍。

### 2.1 基于特征的零件信息模型

早期的 CAD (计算机辅助设计) 系统关心的是零件的几何形状信息, 常以线框模型、面模型或实体模型来描述零件提供的零件信息是几何信息和图形拓扑数据, 对于零件的可加工性考虑很少, 模型中更不表达零件的加工信息, 这种设计方法显然不能满足并行设计的要求。自 1980 年以来, 出现了 AD2000、BDS/GLIDE、BUILD-1、COMPAC、EUCLIDCF 等三维造型系统。当前的 CAD 系统已经达到了相当高的水平, 像 I-DEAS 软件的新版本 Master Series、UCLID-3 以及 Pro/E 等, 都具备强大的造型功能, 简便的操作以及系统内部较完善的信息管理手段, 但这些系统所建立的零件信息模型都还只限于几何信息和拓扑信息, 很少表达面向加工的信息。但如果考虑并行设计中 CAPP 和 DFM 等设计活动对零件信息模型的要求, 则不但要表达零件的几何、拓扑信息, 还要表达有关零件的工程语义信息(表达工程师的设计及制造意图)。为了解决以上问题, 需要开发基于特征的零件信息建模系统, 将零件模型中要求的高层次信息以“特征”的形式表达。

一个完整的零件信息模型不仅是零件产品数据的集合, 还应该表达各类数据的组织方式以及它们之间的联系。当然, 由于信息的组织和零件产品信息的处理方式紧密相关(例如, 面向功能的组织和面向对象的组织方式二者就有很大的差别), 所以, 企图在物理上找出一个统一的零件信息组织是不太现实的。根据计算机系统集成的需要, 以及零件产品生命周期最常用的信息, 零件信息模型需要包括管理信息和制造信息。

零件的功能是靠零件的结构形状和精度来保证的, 而零件的结构形状一般都对应一种或多种加工方

法,每种加工方法又与加工设备密切相关,这样零件功能——结构形状——加工工艺——加工设备就联系起来,而联系的“纽带”,就是“特征”,所以,在设计阶段以特征为基础设计零件就能保证零件的可加工性,“特征造型系统”就是基于这个思想发展起来的。

图 39.5-4 表示了构成零件信息模型的两个层次,

即应用层、逻辑层和表达层。应用层指从用户的角度分析零件对象的属性、功能等,对于表达零件的信息进行抽象、分类;逻辑层是分析满足应用层的逻辑结构表达,以特征的形式表示;表达层就是满足逻辑层各类特征的物理实现方法,用几何/拓扑以及相关信

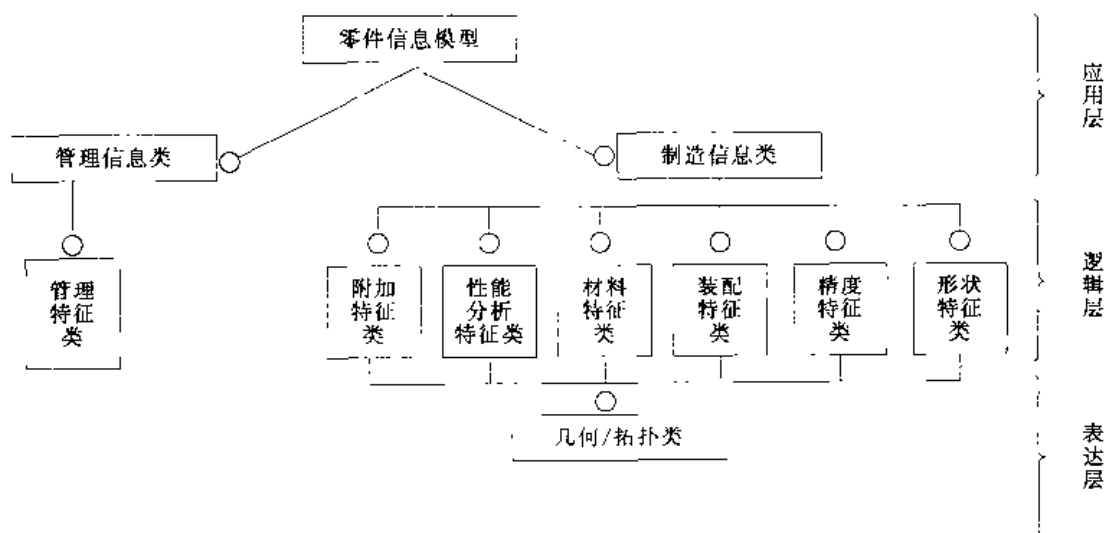


图 39.5-4 零件的特征信息模型

### (1) 特征的作用

1) 充当产品信息描述中信息的组成单元,以表达具有特定功能和特定意义的形状结构。

2) 表达某一个(一组)几何元素或某几个(几组)几何元素之间的加工或工艺属性,如表面粗糙度、形位公差、尺寸公差等。

3) 隐含加工该特征所需要的一道或多道工序。

4) 描述总体信息,以表达零件的属性。

### (2) 图 39.5-4 中列出的各类特征的含义

1) 形状特征类(Form Feature Class) 用于描述有一定工程意义的几何形状信息,是精度特征和材料特征的载体。

2) 精度特征类(Precision Feature Class) 用于描述几何形状和尺寸的许可变动量或误差,例如尺寸公差、几何公差(形位公差)、表面粗糙度等。

3) 装配特征类(Assembly Feature Class) 用于表达零件在装配过程中需使用的信息。

4) 材料特征类(Material Feature Class) 用于描述材料的类型与性能及热处理等信息。

5) 性能分析特征类(Analysis Feature Class) 用于表达零件在性能分析时所使用的信息,如有限元网格划分等。

6) 附加特征类(Additional Feature Class) 根据需要,用于表达一些与上述特征无关的其他信息。

7) 管理特征类(Management Feature Class) 用

于产品的管理信息,是零件的宏观属性描述,如名称、数量等。

在所有的特征类中,形状特征是最重要、最基础的特征,它往往是其他特征的载体。形状特征的分类在不同的应用领域有不同的分类方式。为了实现 CAD/CAPP 的集成一般常常按照零件的结构形状和加工因素来定义形状特征,由于“结构形状”和“加工因素”都是模糊的概念,不同的领域,不同的设计人员都有不同的解释,这就是形状特征定义没有统一的原因。在面向设计的形状特征定义中,一般从零件的功能要求出发,定义能够满足一定的功能要求的设计特征。在面向加工的特征定义中,特征的确定原则是一个特征对应一个或几个加工工序,这样的特征就可以与计算机辅助工艺规划(CAPP)共享特征信息。按照设计定义特征构造出来的零件模型在与 CAPP 集成时,还需要将设计特征映射成加工特征。所以,为了减少特征在映射过程中的困难和保证信息的完整性,一般采用依据加工要求对特征进行分类,通常可以分为主特征和辅助特征。主特征用来构造零件的基本形状体,辅助特征附着在主特征上,也可以依附在另一个辅助特征上,从而形成零件的复杂结构。

## 2.2 特征信息提取和匹配

零件的特征信息模型的获取,一般有三种途径:一是交互式输入工艺信息;二是从已有的 CAD 信息模

型中用某种方式识别或提取面向加工的高层次特征信息；三是特征参与 CAD 系统本身的造型过程。各种方法的简单介绍如下。

1) 交互式输入工艺信息，目前商品化 CAD/CAM 系统均采用该方式，CAPP 需要人工完成，该方法由于指导规则较少，完全由工艺师的经验而定，所以费时、易出错，但较经济实用。

2) 在商品化 CAD 系统生成信息模型的几何模型的基础上，由计算机抽取特征，并将特征信息与几何信息相匹配，这就是所谓的特征信息提取与匹配。由于几何模型很复杂，在识别过程中存在二义性，现阶段只限于简单特征的识别。

3) 特征参与 CAD 系统本身的造型过程，即用特征进行设计，其功能决定于 CAD 系统的功能及其开放性，不仅要能造出零件形状而且能加入工艺信息，如公差、技术要求等，还必须让用户能加入特征，并能方便地识别。在特征中不仅含有几何信息，还含有其他信息。

在特征信息提取与匹配中，零件模型信息可以分为几何信息和非几何信息，其中几何信息是其他非几何信息表达的基础，如面是组成形状特征的基本元素、中心线或边可以作为尺寸的作用对象等。因此，首先应处理好零件的几何、拓扑信息，建立适当的几何模型，在此基础上才能进行特征信息的提取及匹配工作。

特征信息提取与匹配的结果是生成点、边、环、面、特征等数据表，表中记录了特征信息。然后由这些特征信息，遵照应用协议可生成 STEP 中性文件。

### 2.3 产品可制造性评价

产品可制造性的评价内容主要从技术指标和经济指标，如图 39.5-5。技术指标主要从技术角度满足产品的可制造性，评价技术上的合理性和可行性。在满足功能要求的条件下，保证生产费用最低、生产周期最短。

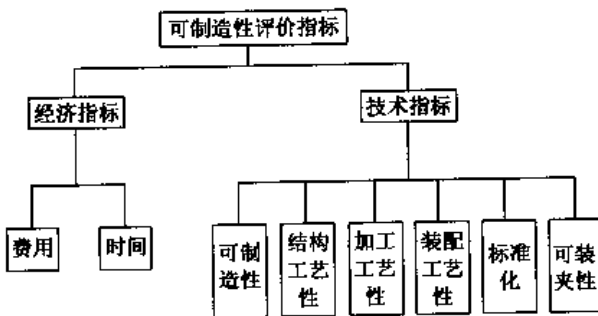


图 39.5-5 产品可制造性评价指标

在并行设计过程中，产品的设计要考虑产品的可制造性，对于产品的可制造性评价贯穿于整个设计过

程。产品的生产环境作为产品设计的约束条件，由于产品的复杂性和产品设计初期的不完整性，为产品的可制造性评价带来许多困难。产品可制造性评价有以下特点：

1) 整体性 由于产品的制造是产品最后一个生产过程，所以产品的方案设计——技术设计——施工设计各个阶段都要考虑产品的可制造性，并且在每个阶段都要实时的进行评价，对产品的设计提出改进意见，产品可制造性的评价贯彻整个设计过程。

2) 针对性 由于在产品设计的不同阶段，例如概念设计、结构设计、详细设计等，设计的阶段不同，技术要求也不同，所以可制造性评价的方法和内容也不同。

3) 模糊性 由于设计的初期信息的不完整性，产品结构和零件的形状难以确定，相应的工艺信息比较缺乏，给产品的可制造性的评价带来相当大的困难，所以早期的评价是不准确的，许多输入条件是估算的或假设的，随着设计的深入，产品结构逐步细化，产品的设计信息、工艺信息逐步完善评价的结果趋于准确。

产品可制造性评价在 DFM 中起着非常重要的作用，但是由于它涉及的因素很大，如何组织相关知识，例如新材料、新工艺、新设备、新的测量方法等直接影响 DFM 的实施。一般评价方法有经验方法、定量评估方法、基于知识的评价等。

经验方法从产品的功能出发，提出一些设计公理指导产品设计。一个好的设计必须符合设计公理，在产品设计和制造的每个阶段，每个决策由公理引导，最终设计出制造性好的产品。由于公理是由实践中抽象而来，不同领域、不同的人员对公理的认识有很大不同，所以实施起来比较困难，但是，该方法在特定产品设计中得到了广泛应用。从制造工艺出发，保证设计出来的产品能够在某种特定的环境条件下，采用某种特定的加工工艺加工出来，例如，面向铸造的设计，面向锻压的设计，面向加工中心的设计等，该方法针对性强，一般在产品具有明确的生产单位时经常采用。

定量评估法从缩短生产成本出发，该方法以费用作为评价指标，综合考虑材料费用、加工费用、运输费用、管理费用等，把这些因素用经验公式进行量化处理，近似估计产品总费用。该方法可以利用计算机技术、信息技术，建立材料库、刀具库、机床参数库、生产工艺库等，根据产品的特点，估算加工费用。该方法经常应用于产品的技术报价系统，用于产品的成本估计。从生产工艺出发，该方法以加工方法以及加工方法的难度为基础，结合生产单位的实际情况，对各种加工方法分类，并用百分制表达加工难度和加工费用，找出产品的加工费用，并对不合理的加工方法进行改进，评

估产品的制造性。从产品的性能价格比出发,分析产品的每个功能和每个功能的价值,综合市场需求,精简或增加新的功能,去掉费用大而又增值少的产品的功能。

随着计算机技术和人工智能的发展,产品的制造性评价系统逐渐发展起来,由于各个系统的解决对象不同,系统功能有所差别。

### 3 面向装配的设计 (DFA, Design for Assembly)

装配是产品生产的关键工艺之一,产品的性能和功能是由产品的装配来实现的。零部件的质量再好,若没有好的装配质量,也不会得到合格的产品。并行设计要求在设计阶段考虑到产品的整个生命周期,利用 DFA 技术在概念设计阶段就对产品的可制造性和经济指标进行分析,可以及早发现设计的问题,及早改正设计,避免到后期阶段发现设计错误引起的大修改,可以大幅度地降低产品成本,提高产品的质量和可靠性,缩短产品的上市时间。

DFA 实施的主要目的是根据产品结构图和装配顺序实施对装配结构的分析、评价,对于设计提出改进建议,减少不必要的零件,优化零件结构,使其便于加工、便于装配,减少加工和装配的时间。为提高产品设计质量,缩短产品设计周期,降低产品设计成本打好基础。

DFA 的核心是建立一个产品的装配模型,它包括了产品的零件和部件之间的层次关系、装配关系以及设计参数之间的约束和依存关系。层次关系描述了机械产品的组成关系。产品分解为部件,部件分解为子部件或零件。从装配次序看,就是由零件组装成装配体,

再参与整机的装配。这种层次关系可直接表示为树结构。装配关系反映了零部件之间的继承和约束关系,主要包括了几何关系、连接关系和运动关系。几何关系有平面贴合、平面对齐和相切等类型;联接关系包括螺纹联结、键、联轴器等传动类型;参数约束关系描述了设计参数在层次关系上的继承和约束要求,它的传播和修改直接影响产品的结构和功能。

#### 3.1 产品结构树

产品装配模型的生成是从分析用户需求开始的,产品的设计由面向用户的功能要求开始,通过 QFD 将用户的需求经过逐步分解得到产品功能配置。功能分解法是产品设计中常见的系统设计方法,产品的功能可以分解成若干子功能,一直分解下去直到功能元,从而生成产品的功能树,如图 39.5-6 所示。

而每一个功能或功能元可以用相应的零部件来体现,所以与功能树相对应的应该有一个产品结构树,也就是产品设计功能模型映射成产品树,如图 39.5-7 所示。层次关系的产品结构树表达了产品的组成以及零部件之间的关系,这种层次关系表达了设计者的设计思想,符合功能分析的习惯,零件组成部件,部件组成产品,但由于功能分解的多样性和功能模型到产品结构映射的多样性,这种产品的层次结构为装配工艺和装配顺序规划提供了基本的参考,不能认为这就是合理的装配顺序,应该以装配层次树为基础,建立新的装配信息模型。在装配信息模型中应该包括以下信息:管理信息(一般由产品结构树提供)、零部件之间的装配关系、几何拓扑信息(一般由 CAD 模型提供)。

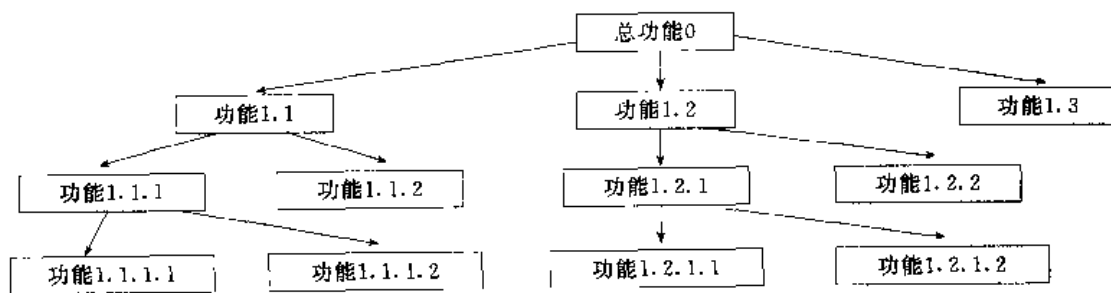


图 39.5-6 产品功能树

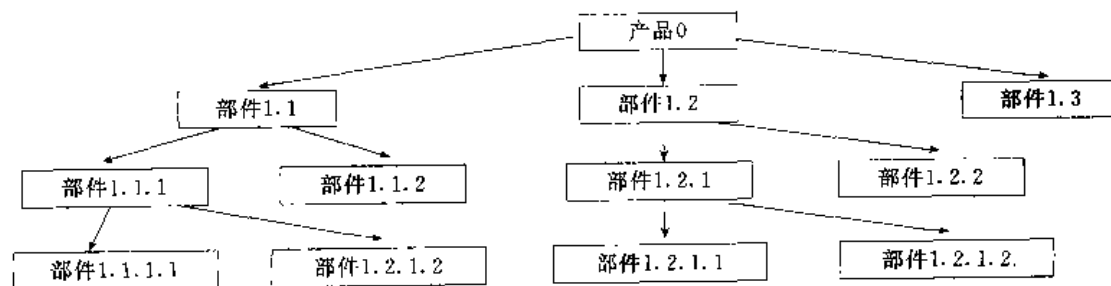


图 39.5-7 产品结构树

### 3.2 装配模型

装配模型是实现 DFA 的基础, 装配模型的先进性和完备性直接影响装配实施的自动化程度和准确性, 建立产品装配模型的目的是为面向装配的产品设计提供信息来源和存取机制。装配模型来自概念设计和结构设计, 通过设计人员的分析给出装配模型和零件的描述(包括: 零部件尺寸、主要结构、位置参数、重量等), 以及零部件之间的连接方式(包括: 连接方式、定位方向), 通过分析整理, 自动生成用于 DFA 的装配模型。也可以利用 CAD 系统生成的部分装配模型。装配模型不仅要包含设计系统的输入信息, 还应能包含设计过程的中间信息和结果信息, 因此装配模型信息应随设计过程的推进而逐渐丰富和完善。由于装配过程的复杂性和信息的模糊性, 至今没有一个满意的装配模型, 而在 DFA 的研究中, 从不同角度出發, 许多研究者提出了自己的装配模型。一般来说, 在装配模型中应包含如下内容:

1) 产品定义 即产品及其零部件管理相关的信息。包括产品各构成元件的名称、代号、材料、件数、技术规范或标准、技术要求, 以及设计者或供应商、信息版本等。它们是在产品设计过程中逐渐形成的, 主要作用是产品为产品设计过程以及产品生命周期后续过程的管理提供参考和基本依据。

2) 产品结构树 表达产品的层次结构。

3) 产品形状定义 即与产品的几何实体构造相关的信息。它们决定装配元件和整个产品装配体的几何形状与尺寸大小。产品装配模型所需的几何构造信息可直接从相关的产品造型数据库提取。

4) 装配工艺信息 包括各装配元件的装配顺序、装配路径, 以及装配工位的安排与调整、装配夹具的利用、装配工具的介入、操作和退出等信息。

5) 装配关联图 产品装配关系中, 装配单元之间的装配关系可以归纳为定位关系、联接关系和运动关系三大类。通过装配关系图描述装配单元之间的这些关系。装配关联图是描述装配模型的一种方法, 从分析产品装配体的特性出发, 表达零件、零件的属性、零件之间的装配关系以及装配方法。

### 3.3 可装配性评价

根据装配模型, 在装配顺序的约束下, 对产品的可装配性进行分析和评价, 主要从技术特性、经济特性和社会特性三方面进行, 如图 39.5-8 所示。

技术特性即按照装配顺序装配时, 从取得装配元件后根据它的定位、连接等, 到装配到位为止, 在

保证装配精度和功能的条件下, 装配的难易程度。经济特性即根据装配顺序, 对装配单元/零件的材料、标准化程度及耗费的工时等进行评价。社会特性主要是对装配单元的再利用和环境保护等进行评价。可装配性评价的输出是经过评价的装配工艺。

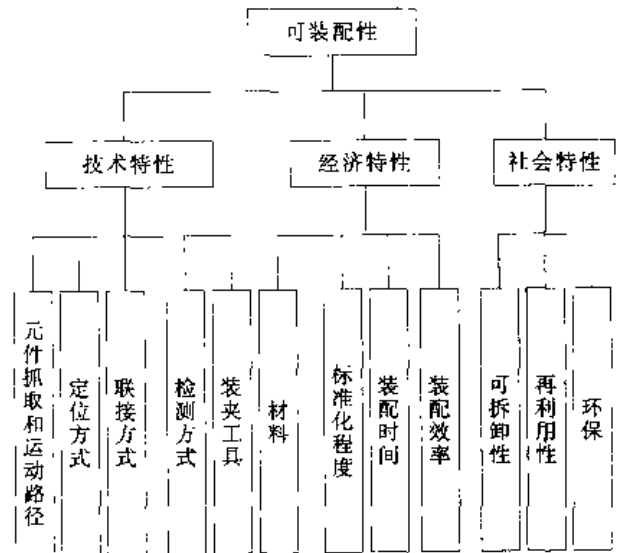


图 39.5-8 产品可装配性评价

## 4 DFM 技术在摆线针轮减速器设计上的应用

### 4.1 摆线针轮减速器设计应用 DFM 技术系统结构

摆线针轮减速器具有传动比范围大、同时啮合齿数多、传动效率高、使用寿命长、结构紧凑等特点, 得到了许多工业部门的应用。以单级摆线针轮减速器为对象, 研究产品 DFM 的系统结构(见图 39.5-9)。根据设计要求确定减速器结构型式、产品主要参数和主要零部件之间的装配关系。在产品主要结构和参数约束下, 对各个主要零件进行设计, 运用特征造型技术, 构造零件的主要几何形状, 并描述零件的加工信息。当各个零件的几何参数确定后, 结合装配关系, 即可进行装配仿真, 确定装配工艺计划; 通过装配评价, 提出对零件间装配关系的调整和零件特征参数的修改。对合适的几个方案可进行零件特征工艺的设计, 根据加工方法库、切削参数库确定各零件加工的工艺路线, 通过可制造性评估, 决定零件的加工工艺方案。由产品的可装配性评价和零件的可制造性评价的综合评估, 确定选用的设计。

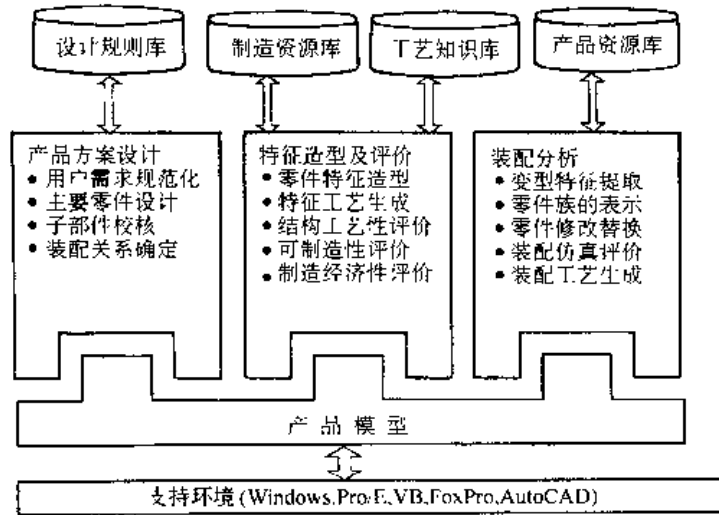


图 39.5-9 摆线针轮减速器应用 DFM 技术的系统结构

## 4.2 产品总体设计

产品总体设计包括产品总体辅助设计模块和产品装配关系确定模块。前者是由设计要求如传动结构型式、传动功率、传动比、使用寿命和输入轴转速（或输出扭矩和转速），来确定产品的主要尺寸。其中，转臂轴承的选择包括初定摆线轮宽度和内孔直径、球轴承的动载荷再选择轴承型号。确定针齿尺寸，首先要确定是否抽齿，确定针齿套直径，检验摆线轮齿廓是否产生根切和尖角，再求针齿销直径。摆线轮结构尺寸包括齿顶圆和齿根圆直径、齿高、齿侧啮合间隙、摆线轮宽度和内孔直径、柱销孔数目和中心圆直径。输出机构设计要确定柱销、柱销套和柱销孔的直径，并验算柱销套和柱销孔的接触强度。产品的传动型式和以上各主要参数存入产品的数据库，以产品 ID 为主索引。这些参数是后续零件设计的约束。

产品装配关系确定模块是结合上述传动型式和主要参数，进一步确定产品的层次关系，明确装配关系和装配约束类型，最后确定零件间的参数约束关系，如输出轴的两支撑点处轴径分别等于两轴承内径等关系。

## 4.3 零件特征造型

零件造型旨在描述零件的形状特征和加工信息。根据产品特点，零件设计通过轴类和盘套类进行造型。如轴类零件的信息包括特征序号、特征名称、特征性质、特征基准点坐标、轴段长度、轴段半径、粗糙度、形位公差和毛坯类型等。这些信息均记录在零件数据库中。对这些特征还可进行复制、移动、参数修改，以及特征类型的改变等操作，这些修改同步地修改了零件数据库中的数据。

## 4.4 装配分析

装配分析包括装配仿真、装配工艺计划确定和可装配性评价等功能。各设计零件按装配关系，传动结构型式可以组装成产品，这时可进行静态干涉检验，并可反馈到零件特征造型模块进行零件参数修改。当描述零件间装配关系和装配约束类型由图结构转变为树结构时，可求出产品的拆卸方案，再反求得到装配工艺方案。由该工艺方案可进行装配的动态仿真和可装配性评价等活动。影响可装配性的因素有部件因素、系统因素和工艺因素。可装配性反映了装配的难易程度、装配时间和装配成本等，其中装配难易程度表示为装配过程中一系列运动的成功概率的乘积。

## 4.5 零件 DFM 评价

由零件的加工特征可按加工方法库确定零件的可选加工方法，结合机床库、刀具库、切削参数库和加工余量库，产生零件的几种加工方案。产品的 DFM 评价就是从结构工艺性分析、可制造性和制造经济性三个方面对产品进行综合评价，提出评价结果、选用方案和修改建议。

结构工艺评价内容有：轴类零件的各段直径应向一个方向或向中间递增；中心孔、进刀槽、倒角、过渡圆角的设置是否合理；避免出现深孔加工。

特征可制造性评价是检查零件的各特征是否符合当前生产环境下的加工工艺、设备资源和工装约束使零件具有较好的可制造性。原型系统通过检验各制造特征的类型及其属性值与相应的机床和刀具的工作参数进行比较，作出可制造性建议。

制造经济性评价则根据切削参数库和加工余量库

计算零件加工的时间和成本,有利于设计者在评价时期就能较好地把握制造成本。

产品评价的基础工作是建库,使建立的各数据库可查阅、可维护,并有修改权限等功能。轴的加工规则库中描述了表面、材料、尺寸上下限、公差等级上下限、粗糙度和毛坯等要求,以及该要求下的加工次序和方法。机床库按机床类型分别建立,如车床库描述了车床

ID、机床类型、功率、加工的零件最大长度、零件的最大外圆直径以及空闲状态。机床的空闲状态是在可制造性评价前对制造制源的一次动态确认。刀具库也按刀具类型分别建立。加工余量库包括加工操作性质、材料、热处理、孔径和余量分配顺序等。

经过可装配性评价和可制造性评价的综合分析,确定选用的产品设计方案。



## 第6章 并行设计的集成平台 PDM

### 1 PDM 概述

#### 1.1 PDM 的产生背景

随着计算机技术和信息技术在工业界的应用,企业就不断利用这些高新技术提高自身的设计能力和管理水平,从最早的 CAD、CAM、CAPP 等到现代的 CIMS,从传统的串行设计到并行设计,从 MRP、MRPII 到 ERP 等。在应用的过程中,许多企业发现他们正被大量的产品数据所包围,各个自动化孤岛之间的信息不能充分共享,产品设计过程的信息无法交流。之所以受到这些阻碍,是因为企业在信息化过程中存在以下六点现实问题:

1) 信息共享程度低 企业的计算机辅助工具大都是一些离散孤立的系统,由这些软件产生的数据通常互不兼容,即这些数据常常以不同的格式和介质存储,从而无法在设计、工艺和制造部门之间有效地实现信息共享与传递。

2) 信息传递速度慢 大部分企业的信息传递主要是通过书面文件、报表及电话联络方式,无法及时收集工艺和制造过程中的反馈信息,不能实时反映设计过程中的变化,更谈不上随时跟踪整个产品的设计和制造的进展情况。

3) 业务管理落后 虽然企业采用计算机辅助工具在一定程度上实现了绘图和设计的计算机化,但是有关产品的其他技术资料依然采用手工管理的方式。另外,由于 CAD/CAM 技术推广应用,各种格式的图形、数据、文本等文件还是沿用过去对纸质文件的管理方法,对文件缺乏有效的版本管理,造成不必要的重复设计。由于缺少迅速有效地检索产品数据文件的手段,也造成设计的复用性差。

4) 设计方式陈旧 长期以来,产品开发沿用传统的顺序工程方法,遵循“概念设计→详细设计→过程设计→加工制造→实验验证→设计修改”的流程。由于该方法在设计的前期不能全面地考虑设计下游的可制造性、可装配性、质量保证等多种因素,使所制造的产品存在许多缺陷,这就必然要求对设计进行更改,构成了从概念设计到设计修改的大循环,而且可能在不同的环节多次重复这一过程,造成设计改动量大、产品开发周期长、成本高,难以满足激烈的市场竞争的需求。

5) 支撑技术不配套 在数据管理方面,企业仍然停留在使用大型商用关系数据库(如 ORACLE、SYBASE、INGRES 等)的层次上,由于这些数据库不能有效地管理图形、图像等非结构数据,更无法实现过程管理、配置管理以及对应用工具的集成,从而不能满足在异构与分布式计算机环境中使企业内各类应用实现信息集成、功能集成和过程集成的目标。

6) 应用集成系统效率不高 不同的应用软件目前只能做到几何/拓扑层信息的直接分享和交换,其他诸如特征、操作层信息还难以直接交换。

为了解决这些问题,产品数据管理(Product Data Management,简称 PDM)应运而生。

#### 1.2 PDM 的概念

PDM 是以软件为基础,管理与产品相关的信息(包括电子文档、数字化文件、数据库记录等)和所有与产品相关的过程(包括审批/发放过程、工程更改过程、一般工作流程等)的技术。它提供产品全生命周期(包括市场需求调研、产品开发、产品设计、销售、售后服务)的信息管理,并可在企业范围内为产品设计和制造建立一个并行化的协作环境。

PDM(Product Data Management)的概念起源于 20 世纪 80 年代末期,当时的目的只是为了解决大量工程图样文档的管理困境,随着时间的推移,其应用逐渐扩展到产品开发整个生命周期的数据管理问题。到目前为止,对于 PDM 尚无一个确切而又完整的定义。在 1995 年 2 月,主要是致力于 PDM 技术和相关计算机集成技术的国际权威咨询公司 CIMdata 公司总裁 Ed Miller 在《PDM Today》一文中给出了简单定义:PDM 是一种用来管理与产品相关的所有信息(包括零件信息、产品配置、文档、CAD 文件、审批信息等)和与产品相关的所有过程(包括设计的组织者、设计人员、信息的控制、权限的审批和分发等)的技术。而 Gartner Group 公司的 D. Burdick 在 1995 年 9 月所作的《CIM 策略分析报告》中,把 PDM 定义为:PDM 是在企业范围内为设计和生产构筑一个并行化产品协作环境(由供应、工程设计、制造、采购、销售、市场、客户构成)的关键使能技术。

尽管 PDM 的定义不尽相同,至今也没有一个统一的认识和完整的定义。但大体上可分为广义和狭义

两种定义。从狭义上讲，PDM 仅管理与工程设计相关的领域内的信息；从广义上讲，它应该能覆盖到整个企业中从产品的市场需求、研究开发、产品设计、工程制造、销售、服务、维护等生命周期中的信息。

总结各种不同的定义，可以得到这样的理解：从技术的角度，产品数据管理 (PDM) 是一种管理技术，管理所有与产品相关的信息和过程的技术；从软件系统的角度，PDM 是一种管理软件，提供帮助工程师以及其他人员管理产品资料与开发步骤的一种软件系统，提供数据、文件、文档的更改管理、版本管理、产品结构管理和工作流程管理等基本功能。从工具的角度，PDM 是在数据库基础上的软件技术，是介于数据库和应用软件之间的软件开发平台，可以集成或封装多种开发环境和工具，是企业全局信息集成的理想平台；是支持企业过程重组、实施并行工程和 CIMS 工程的使能技术；同时也是一个面向对象的电子资料室，能够集成产品生命周期内的全部信息。

综合上述定义，PDM 可以定义为以软件技术为基础，以产品为核心，实现对产品相关的数据、过程、资源一体化集成管理的技术。PDM 明确定位为面向制造企业，以产品为管理的核心，以数据、过程和资源为管理信息的三大要素。PDM 进行信息管理的两条主线是静态的产品结构和动态的产品设计流程，所有的信息组织和资源管理都是围绕产品设计展开的。

### 1.3 PDM 的特点

虽然对于 PDM 还没有一个统一的定义，尽管它们在实施技术和功能上有一些差别，但是，目前的商用

PDM 系统在组织方式上大体是一致的，在体系结构上是类似的，有如下共同特点：

采用面向对象的开放体系结构，为用户根据本企业的情况本地化、系统功能扩充提供支持，并对多种应用软件留有标准接口；

采用客户/服务器的结构，为实现多用户协同工作提供一个方便灵活的运行方式；

为用户提供开发接口，从而为进行二次开发或集成其他应用软件打好基础；

支持多种数据库存储，如分布式数据存储、元数据存储；

支持多种网络通信协议，如 TCP/IP、Netware 等；

支持不同的硬件平台：如 RS/6000、SUR、SGI 等；

支持多操作系统：如 UNIX、SUN-OS、Windows NT 等；

与该 PDM 系统集成良好的 CAD/CAM 软件如 IMAM 系统与 UGII 软件，WINDCHILL 系统与 Pro/E 软件。

### 1.4 PDM 的一般体系结构

PDM 系统应用面向对象的设计方法，采用现在流行的 Client/Server 体系结构，并建立在分布式数据库技术和网络技术的基础上。PDM 的结构大致可分为四个层次：用户层、应用功能层、应用服务层和系统服务层，如图 39.6-1 所示。

用户层：用户层主要是给用户提供了 PDM 的功能描述和友好的用户界面。该界面是由 PDM 系统依据功能单元的不同和用户的账号动态分配的，这样就

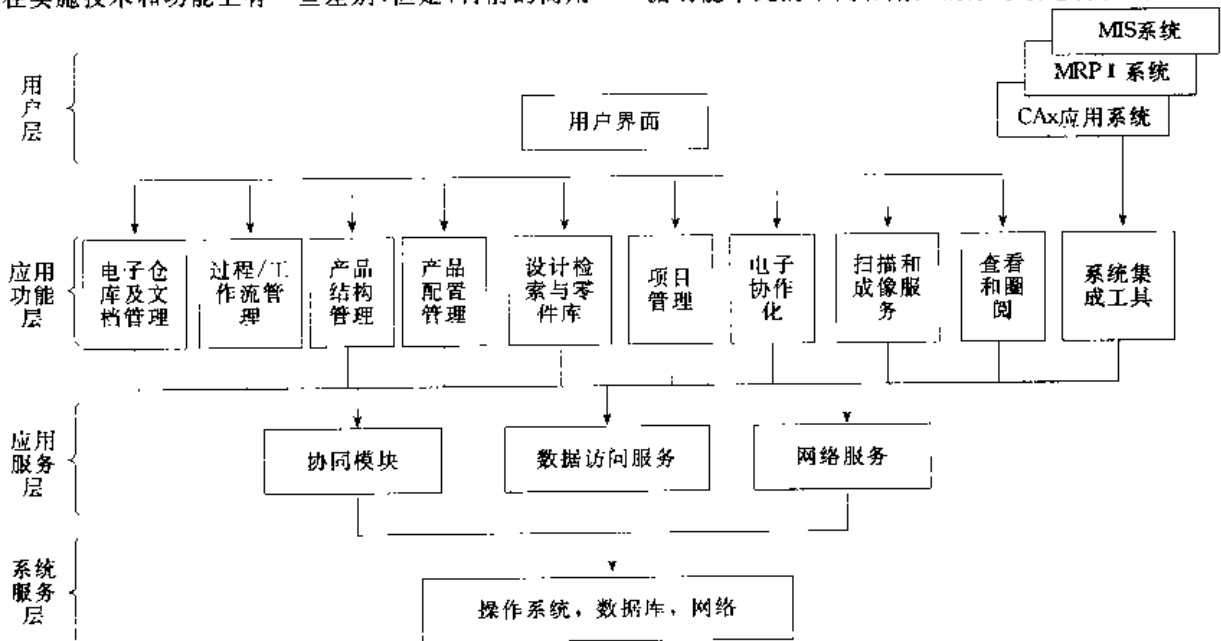


图 39.6-1 PDM 的一般体系结构

使得 PDM 系统只展现该用户所关心的产品信息。

**应用功能层：**应用功能层包括 PDM 软件的一系列功能模块，其中主要有电子仓库和文档管理、过程/ workflow 管理、产品结构管理、产品配置管理等模块。这些模块都是 PDM 系统中的重要组成部分，完成产品数据管理的大部分工作。系统集成工具也包括在这一层，它是 PDM 与 CAX、MRPII、MIS 集成的重要开发工具。

**应用服务层：**应用服务层主要是为应用功能层中的功能模块提供服务，也可以说是应用功能层与系统服务层之间的桥梁。该层的功能是协调各功能模块之间的关系，处理功能模块对数据库和网络的访问。

**系统服务层：**这一层主要是包括操作系统、数据库系统以及网络运行环境。这一部分是 PDM 系统实施的基础。

## 2 PDM 的主要功能

PDM 系统的功能按照软件模块来划分，可以概括为八个方面：

(1) 电子数据仓库 (Vault) 及文档管理 (Document Management)

电子数据仓库是 PDM 最基本的功能，通过电子数据仓库，可以生成、存储、查询数据等，还控制存取、恢复、编辑电子文档等。不同用户或角色按照操作权限，对电子数据仓库中的数据分别进行操作。电子数据仓库提供了存取产品生命周期内所有数据的能力，电子数据仓库中还保存了管理数据的数据（称为元数据），这些数据是指向描述产品的物理数据和文件指针，为 PDM 控制环境和管理提供依据。借助于分布式数据库技术，它还允许用户迅速地访问其权限内全企业的所有产品信息，而不必考虑用户和数据的物理位置。

文档管理提供对电子数据仓库中的文档的安全控制与管理，如对象归档、版本控制、签入、签出、权限控制、对象状态的改变、对象的查询、动态浏览与导航等。

电子数据仓库和文档管理能提供多用户的并行访问。为了防止两个用户同时修改同一个数据，系统采用 Check-in 和 Check-out 机制。当一个用户打开某个数据进行编辑时，系统自动对该数据“写”访问权进行加锁控制，任何其他人都不能对它进行修改。只有当原先打开该数据进行编辑的用户将该数据重新放回数据库或关闭该数据时，该数据才恢复原来状态，即允许其他授权用户进行修改。手工的 check-Out 命令可以对数据设置一个长时间的锁，防止该数据发生变化；Check-in 则用来解除原先设置的锁使数据恢复到可修改的状

态。

(2) 工作流和流程管理 (Workflow and Process Management)

工作流和流程管理主要是用来定义和控制数据操作的基本过程，通过创建、执行、监督和控制工作流程，可以满足用户在开发和制造产品过程中的各种操作和执行过程的需要。用户可以建立串行或并行的工作流程，使得工程技术人员并行地协调工作。工作流和流程管理的原则就是围绕设计对象定义其工作流程，并定义所提交产品数据的版本和修改权限控制。工作流和流程管理可以根据所定义的工作流程、工作人员和工作权限及时地通知每一个人，每个用户在登录后，检查自己的任务表，就能清楚地知道自己应做的工作和完成期限。同时，项目负责人可以随时了解项目的进展情况。在整个设计过程中，产品的设计和每一次修改都通过历史版本记录而保存下来，工程人员可以根据记录查询以前所做的所有工作，以便进行更好的设计。

(3) 产品结构管理 (Product Structure Management)

产品结构管理 (PSM) 提供管理产品的结构定义和配置，允许用户定义所有与产品相关的数据。同时，产品结构管理还能建立数据与数据之间的关联，提供建立、编辑产品结构树的基本功能，并能控制相关数据的修改，从而保证产品数据的移植性与完整性。

对于产品生命周期中不同阶段，如“设计阶段”，“组装阶段”等，通过产品的结构配置，提供不同的产品结构视图，满足不同阶段的用户所需的产品结构视图需求。

实现产品结构配置功能通常有三种方法：产品版本配置 (Revision Configuration)、时效配置 (Effectivity Configuration) 和变型配置 (Variant Configuration)。

(4) 产品配置管理 (Product Configuration Management)

产品配置管理以电子数据仓库为基础，以系列化产品的信息相关性作为管理对象，以材料清单为核心，根据产品结构及产品系列化的要求，进行产品信息的归类，实现产品数据的组织、控制和管理。产品配置管理一方面继承了产品结构管理的全部功能，同时增加了配置项、结构选项（互换件、替换件、供应商）。通过提供结构有效性与版本有效性进行管理，描述更为复杂的产品结构关系。

(5) 设计检索与零件库 (Design Retrieval/Component Libraries)

设计检索和零件库功能提供了通过分类和编码来有效地组织产品数据。零件库是利用成组技术将具有

相同或相似特征的零件归并入库,以更好支持原有设计资源的重新利用,从而提高设计效率。设计检索则可以通过多种检索途径来获取产品零部件及其相关的设计数据。PDM 软件能够提供分类规则的定义方法,以适应每个企业的信息管理要求,同时提供有力的检索工具,可以根据产品对象的名称、属性、编码等进行产品信息的查询,并能查询零件的各种信息,如状态信息、创建者信息、被哪些产品所使用、或零件的使用频率等。设计检索与零件库访问是在开发新产品时重新利用已有的设计信息及提高产品标准化的关键所在。

#### (6) 项目管理 (Project Management)

项目管理是在项目实施过程中实现其计划、组织、人员及相关数据的管理与配置,能够为管理者提供项目的实时活动和状态信息,完成计划的反馈。目前的 PDM 提供项目管理功能很弱,许多 PDM 产品只能提供 workflow 管理。通过 PDM 与流行的项目管理软件封装,可以完成计划/项目的管理功能。

#### (7) 系统集成 (Integration)

为了使不同应用系统之间能够共享信息以及对应用系统所产生的数据进行统一管理,要求把外部应用系统“封装”和集成到 PDM 系统中,并提供应用系统与数据库以及应用系统与应用系统之间的信息集成。这就要求 PDM 系统必须具有强大的客户化和二次开发能力。现在大多数 PDM 产品都提供了二次开发工具包, PDM 实施人员或用户可以利用这类工具包来进行针对企业具体情况的定制工作。

#### (8) 其他功能

除了以上七类功能以外, PDM 可以提供如电子协作功能 (Electronic Collaboration)、扫描和成像服务 (Scanning & Image)、查看和圈阅功能 (View & Markup) 等功能。

### 3 PDM 的集成

并行设计过程中,涉及到许多软件工具,分别用于设计、测试、分析、仿真等,这些工具既是数据的使用者,也是数据的产生者。为了便于数据的管理,保证数据的一致性和完整性,必须把这些系统集成在 PDM 管理框架中,由管理框架对数据进行统一的控制和管理。PDM 系统与其他系统的集成有利于数据在不同应用领域的重复使用,增进各部门之间的交流与合作,消除冗余和手工出错,提高效率。

#### 3.1 PDM 集成的种类

根据 PDM 系统本身在集成中所处的地位, PDM 系统与其他应用系统的集成有如下三种关系:

- 1) 从属集成关系 PDM 系统本身作为一个子系统被集成;
- 2) 平行集成关系 PDM 系统与其他系统处于平行关系;
- 3) 包容集成关系 PDM 系统作为集成框架,集成其他子系统。

三种集成关系如图 39.6-2 所示。

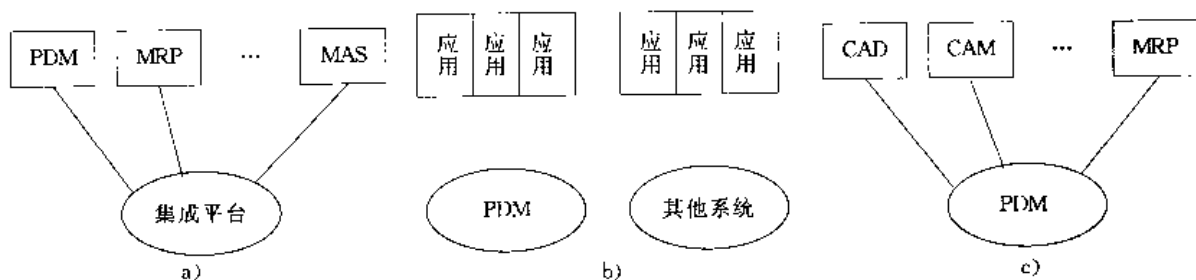


图 39.6-2 PDM 与其他系统集成关系分类

a) 从属集成关系 b) 平行集成关系 c) 包容集成关系

#### 3.2 PDM 与应用系统集成的层次

基于 PDM 框架的应用系统集成可以分为三个层次:接口 (interface) 模式、封装 (encapsulation) 模式与集成 (integration) 模式。基于不同集成层次的企业 PDM 系统集成示意图如图 39.6-3 所示。

##### (1) 封装模式

封装是指把对象的属性和操作方法同时封装在定义对象中,用操作集来描述可见模块的外部接口,从而保证了对象的界面独立于对象的内部表达。封装将对

象封闭保护起来,管理着对象的内部状态,这就意味着用户“看不到”对象的内部结构,但可以通过调用操作即程序部分来使用对象,这充分体现了信息隐蔽原则。由于封装性,当改变一个对象类型的数据结构内部表达时,可以不改变在该对象类型上的任何程序。封装使数据和操作有了统一的模型界面。

封装是对产生数据的应用程序的集成,如 PDM 通过对外部应用软件系统的数据进行封装,使应用软件生成或输入的数据成为 PDM 系统“可识”数据,以便对其进行有效的管理。图 39.6-4 中,在封装模式下,

PDM 系统并不需要了解数据的内部结构,只要在对数据进行编辑时,启动与数据相应的应用系统程序,并输出原始数据,最后关闭应用系统及取回运行结果即可。

而在应用系统对数据进行处理的过程中,PDM 系统无法对应用系统进行控制,应用系统也不能使用 PDM 系统的功能。

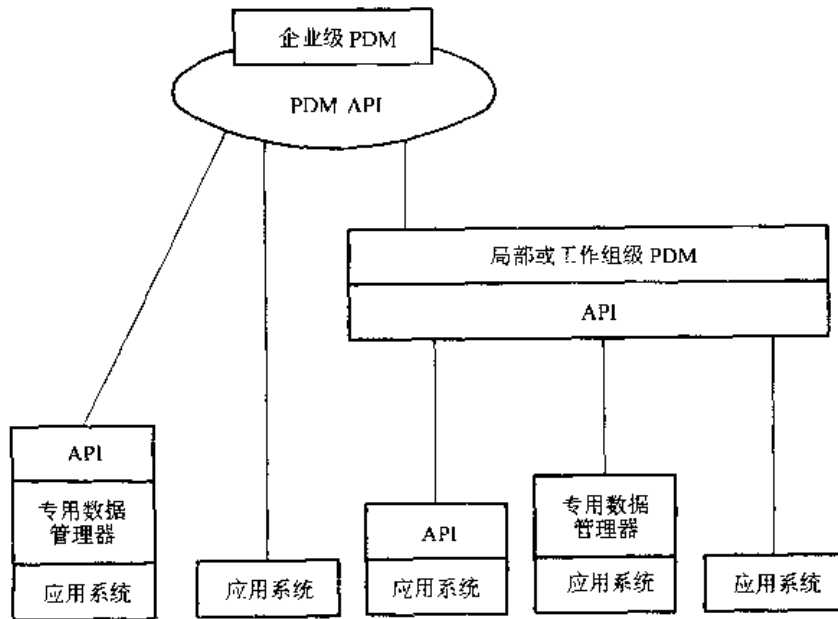


图 39.6-3 PDM 与应用系统间不同层次的集成示意

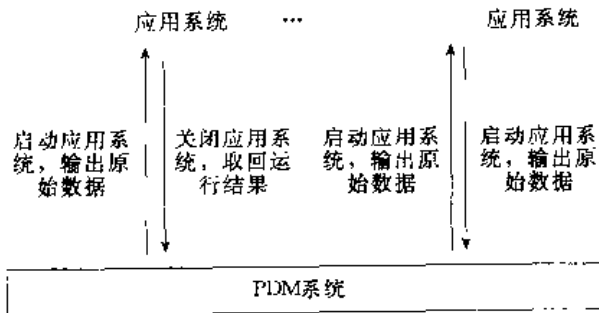


图 39.6-4 数据交换的封装模式

(2) 接口模式

接口是不同系统之间为了保证数据的一致性而进行的一种文件级的数据交换模式。在利用接口进行数据交换时,双方各自是独立的系统,处于平等地位,按照一定的数据转换协议,采用一定的数据格式进行交换。不同的 CAx 系统之间有许多数据交换接口标准,如初始图形信息交换规范 (IGES) 数据传输和交换接口 (SET)、VDA 曲面数据交换格式 (VDA-FS)、CAD 接口及产品数据交换规范 (PDES) 等。在这些规范的基础上,由 ISO 制定了新一代产品数据交换规范 STEP,这个规范的目的是覆盖过去所有的接口标准的功能和应用范围,并解决产品数据交换中存在的问题。接口模式又可分为以下 2 种:

1) 专用数据格式的文件交换产品信息 各应用系统所建立的产品数据模型各不相同,相互间的数据

交换需存在于两个系统之间。其特点是原理简单,转换接口程序易于实现,运行效率较高,见图 39.6-5。但当系统较多时,接口程序增多,编写接口时需要了解的数据结构较多,当一个系统的数据结构发生变化时,引起的修改量也较多。这是 CAD/CAPP/CAM 系统发展初期所采用的产品信息交换模式。这种接口标准如 Autodesk 公司的 DXF、Intergraph 公司的 ISIF 等。

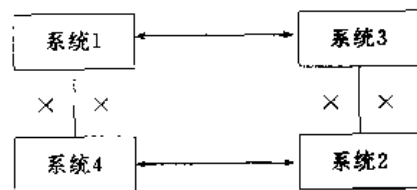


图 39.6-5 数据交换的专用接口模式

2) 标准数据格式的文件交换产品信息 系统中存在一个与各子系统无关的标准格式,各子系统的数通过前置处理转换成标准格式的文件,也可以通过后置处理,将标准格式文件转换为本系统所需要的格式。在这种模式下,每个子系统只与标准格式文件打交道,无需知道别的系统细节,为系统的开发者和使用者提供了较大的方便,并大大减少了系统内的接口数和降低接口维护难度,见图 39.6-6。但这种模式需要解决各子系统间模型统一问题,且运行效率较低,也不能算是一种十分理想的产品信息交换模式。

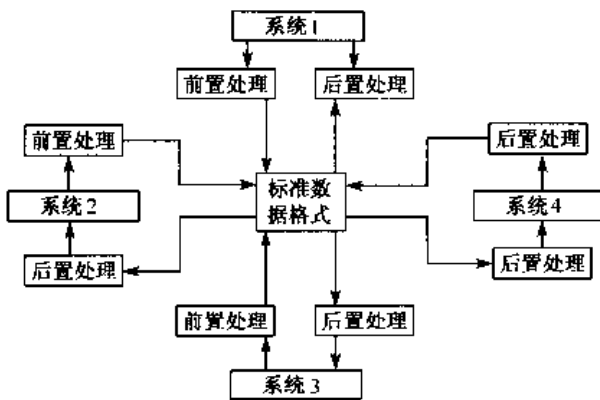


图 39.6-6 数据交换的标准接口模式

### (3) 集成模式

数据交换的集成模式采用有统一的产品数据模型，并采用统一的工程数据库管理系统来管理集成数据，使各系统之间直接进行信息交换，提高数据传输效率和系统的集成化程度，保证各系统之间数据的一致性、可靠性和数据共享，真正实现应用软件系统之间的信息交换与共享。

上面三层集成模式各有利弊，企业可根据实际情况在 PDM 系统中，实现与应用系统之间不同层次的集成。

## 3.3 PDM 对数据信息集成的方法

PDM 是连接设计、工艺、生产的桥梁和纽带，维持着数据的一致和完整。PDM 对数据信息集成的方法有两种：

### (1) PDM 系统内部数据的集成

实现方法是，在 PDM 内部建立一个单一的数据源 (SSPD)，将设计、工艺、生产所需的数据统一保存和维护，避免数据的冗余和混乱，从而达到整个系统数据的共享和集成。建立单一的数据源有如下好处：

1) 数据共享 从用户和应用系统来看，系统的全部数据都存储在 SSPD 中，各种数据只存有惟一的一份，各应用系统所需的数据都从 SSPD 中取得，它们生成的共享信息也都存入 SSPD。这样，各系统只需具有与 SSPD 的接口，即可对信息进行共享，而不必涉及相互之间的复杂关系和数据如何存储，简化了应用系统的设计。

2) 降低了信息集成的难度 系统将全局共享的信息集中到 SSPD 中进行统一的管理，避免了将信息分散到各个应用系统中引起的格式不一致等问题，降低集成的难度。

3) 有利于系统的管理 数据和处理数据的应用软件是分开的，可以方便地采用不同的应用软件数据

处理工作，并且从 SSPD 即可完成全部的数据管理工作，如安全管理、备份恢复等。

(2) PDM 系统与外部系统 (如 CAx、ERP 等) 的数据集成

实现方法是，通过 PDM 所提供的与其他系统的集成接口，通过接口达到系统数据交互使用和模块互操作等功能，从而实现企业信息一体化和集成的目的。系统的集成接口包括数据接口和操作接口两部分：数据接口是为解决数据转换和数据格式不一致的问题，如实现从 EBOM (工程设计 BOM) 到 PBOM (工艺制造 BOM) 的数据转换，以保证系统数据能够形成各种正确的信息流，被系统用户广泛、充分地利用。操作接口的目的在于，使各系统达到融合，让 CAx、ERP 等外部系统的某些注册用户可操作 PDM 系统中的一些功能，同样 PDM 系统中的某些授权用户可操作外部系统中的一些功能。

## 3.4 基于接口模式的应用集成策略

PDM 系统是对产品整个生命周期的数据进行统一管理，它可以将 CAx、ERP 的数据进行集成，并作为连接 CAx 系统和 ERP 系统的桥梁，利用接口使 CAx、ERP 和 PDM 系统进行数据互访，并通过对数据库的访问达到集成的效率，其体系结构如图 39.6-7 所示。

### (1) CAD 与 PDM 的接口

此接口的实现需结合使用 CAD 本身的开发功能和 PDM 的对象服务功能。通过接口，PDM 与 CAD 可实现与工程图相关的管理信息的全自动双向交换，CAD 可自动提取图样明细表中的相关信息。根据其明细表内含有的装配关系，生成 EBOM (工程设计 BOM)，CAD 生成的工程图文件、产品结构模型等都由 PDM 管理。

### (2) CAPP 与 PDM 的接口

此接口的实现需结合 CAPP 的数据访问功能与 PDM 的对象服务功能。CAPP 从 PDM 中获取设计形成的 EBOM，根据实际制造情况进行各种产品零件的工艺路线与工序设计，生成 PBOM (工艺制造 BOM) 产生工序图、工艺文件、工时定额等信息，并作为 PDM 的对象服务存储到 PDM 中。

### (3) CAM 与 PDM 的接口

此接口的实现需结合 CAM 的数据访问功能与 PDM 的对象服务功能。通过接口 CAM 从 PDM 中获取三维图形信息和工艺信息，生成可供数控机床使用的 NC 代码和加工时间等信息存储在 PDM 中，并完成零件的刀具模拟仿真。

### (4) ERP 与 PDM 的接口

ERP 存储生产计划、财务、设备等信息,但所需的一些基础数据必须从 PDM 中获得。加工处理后形成的数据对 PDM 数据模型是有影响的,也必须存入 PDM 系统。ERP 与 PDM 的数据接口有两种,即基本接口和扩展接口。基本接口是 PDM 为 ERP 提供数据

访问服务,ERP 的基础数据访问模块通过基本接口得到它所需的 PBOM 等基础数据。扩展接口是 PDM 为 ERP 中的数据向其他模块提供的访问接口。由于两者的特殊关系,两个系统之间可能要绕过对象层,而直接访问库、表结构。

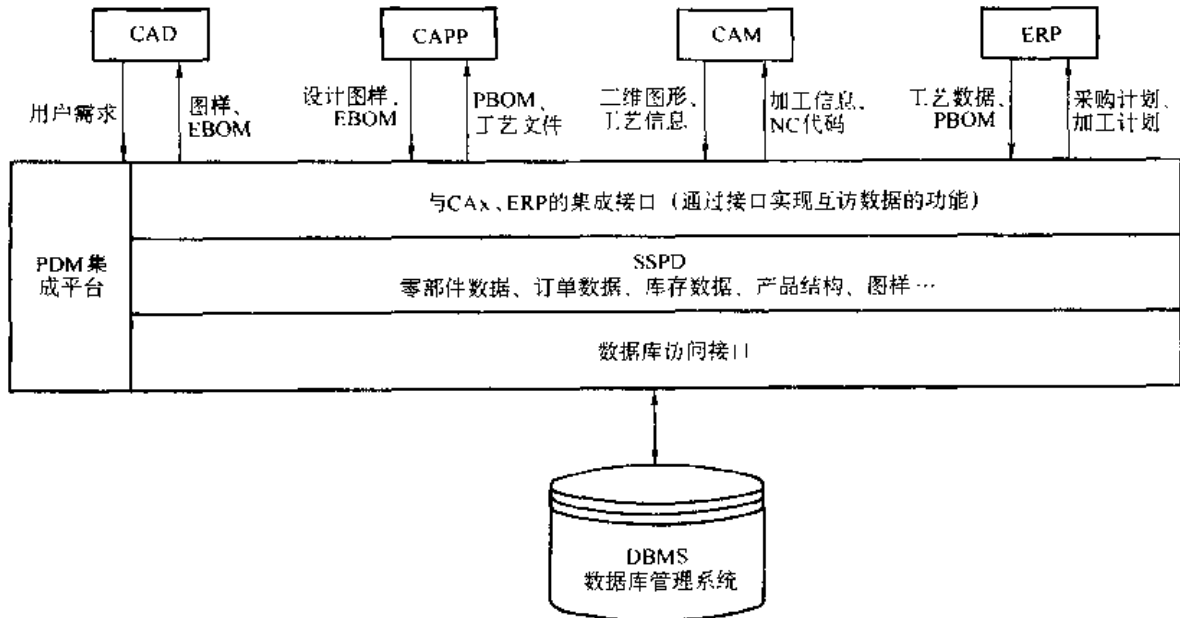


图 39.6-7 PDM 与 CAx、ERP 的应用集成结构

#### 4 PDM 的实施

目前流行的 PDM 软件各自都有各自的特点,功能也有所不同,从功能上可分为三大类:

第一类是以文档管理、数据管理为重点,主要用于文档管理;第二类是以设计过程管理及产品数据管理为重点,主要面向设计过程;最后是面向软硬件系统的集成平台,主要解决企业级的信息集成和过程集成。在选用 PDM 软件时,应该充分注意到 PDM 的应用范围,从企业的实际需要出发,选用合适的 PDM 产品,也可以根据企业的发展规划,选用功能齐全的 PDM 产品,采用统一规划、分步实施的策略。一般来说,PDM 实施可以分为三个层次。

##### (1) 实现文档管理

选用第一类 PDM 就可以满足要求。根据产品的各种数据、图样类型建立电子数据仓库,实现文档分类管理与产品、零件分类管理,方便查询,实现产品资源及信息的重用和共享。

##### (2) 实现产品设计过程的管理

建立基于产品设计的产品模型和产品设计的工作流模型,以新产品设计、系列设计为对象,建立版本管理、权限定义、组织管理以及配置管理规划和结构管理的配置关系等,实现产品配置的多视图管理,支持 CAD、CAPP、MRPII 等与 PDM 系统的集成。在网络技术的支持下,从单一服务器的集成实施走向多服务器的分布式实施。

##### (3) 实现企业级的管理

基于企业的信息模型和过程模型,建立企业内部各种工作流程,优化企业的工作规范,实现企业级的应用系统与 PDM 的紧密集成。

在实施 PDM 的过程之前,首先应了解企业的产品开发过程,全面了解企业的硬软件环境。由于 PDM 的实施,可能引起企业产品开发方法的改进和组织的变更,所以必须探讨企业今后发展的方向,定出与企业发展方向一致的实施方案。

## 第7章 并行设计实例

### 1 基于PDM的油泵产品并行设计原型系统

#### 1.1 系统目标

1) 建立面向并行设计的运行环境。根据并行工程机理,以油泵类产品为背景,以异构网络、分布式数据库和PDM软件为基础,建立面向产品并行开发的运行环境。该并行开发环境支持产品开发人员的分布、协同工作,为产品数据和开发信息的并行操作、管理与控制提供了条件。

2) 建立面向并行工程的产品开发过程管理模式。根据优化的油泵产品开发过程,对产品开发活动实施有效的监控与管理,建立过程的管理模式和冲突的协调方式。

3) 建立支持并行工程的集成产品开发团队(IPT)。在协同开发环境的支持下,建立多层次的油泵产品集成开发团队,并建立IPT的任务指派、管理与运行模式。

4) 建立典型产品零部件实例库。充分利用已有产品实例,提取相似产品的信息,建立油泵类产品重要零部件的实例库,指导产品进行合理的设计。

5) 建立产品设计方案评价及模糊再设计,指导产品的优化设计。

6) 应用工具的集成。主要包括CAD软件与PDM软件的集成、CAD软件的零部件分类工具与PDM软件零件族管理功能的集成、油泵类产品辅助设计工具与PDM软件的集成等。

#### 1.2 体系结构

该体系结构包括四个层次(图39.7-1):

1) 控制层 控制层完成对过程的监控与管理。并行工程的开展是围绕着过程这个核心的,过程和过程中的活动决定了产品数据的产生角色、时间和类型,决定了团队的组织模式、人员的职责和权限、资源的分配等关键问题。控制层的作用就是对过程进行必要的管理,协调矛盾,解决冲突,使并行过程得以顺利地推进。

2) 应用层 应用层面向不同的应用领域,完成不同的任务,如CAD、CAE、CAM、DFA、各种专家系

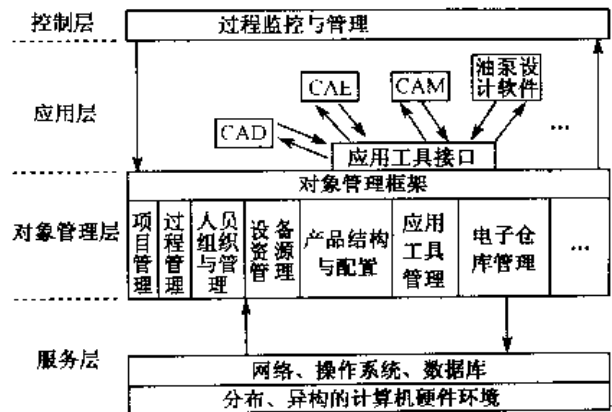


图 39.7-1 以PDM为支持框架的并行设计体系结构等等。该层是数据的产生器和接受器,它通过应用工具接口封装在对象管理框架中,实现“即插即用”。

3) 对象管理层 这一层以 Metaphase 的 OFM (对象管理框架)为核心,实现对并行过程所涉及到的对象进行定义和管理。这些对象包括:项目、过程、人力资源、物力资源、电子仓库、应用工具、产品结构与配置等等。

4) 服务层 服务层为上层提供支持,包括操作系统、数据库和网络。

#### 1.3 IPDT 设计

并行设计强调 IPDT 的组织和管理,一个理想的 IPDT 应由与产品生命周期中各阶段相关的人员组成。IPDT 成员按功能不同分为不同的角色,在系统规划时由系统规划人员根据需要对不同的角色赋予不同的权限、职责和相应的视图。一个角色可以由一个人组成,也可以由多个人组成;同样,一个人可以扮演一个角色,也可以扮演多个角色,这完全可由系统规划人员根据工作量多寡和各人能力高低做适当的调整。合理配备产品开发人员队伍,将产品开发过程模块化、并行化、计算机化。

在油泵类产品研发过程主要涉及设计人员、工艺人员、仿真和配试人员。通过分析,将用户组织成为 13 个角色,即总设计师、柱塞偶件设计师、出油阀设计师、凸轮设计师、油泵组组长、喷油器设计师、喷油器组组长、仿真工程师、仿真组组长、工艺师、配试工程师和配试组组长。它们与活动的关系见油泵并行开发过程图(图39.7-2)。



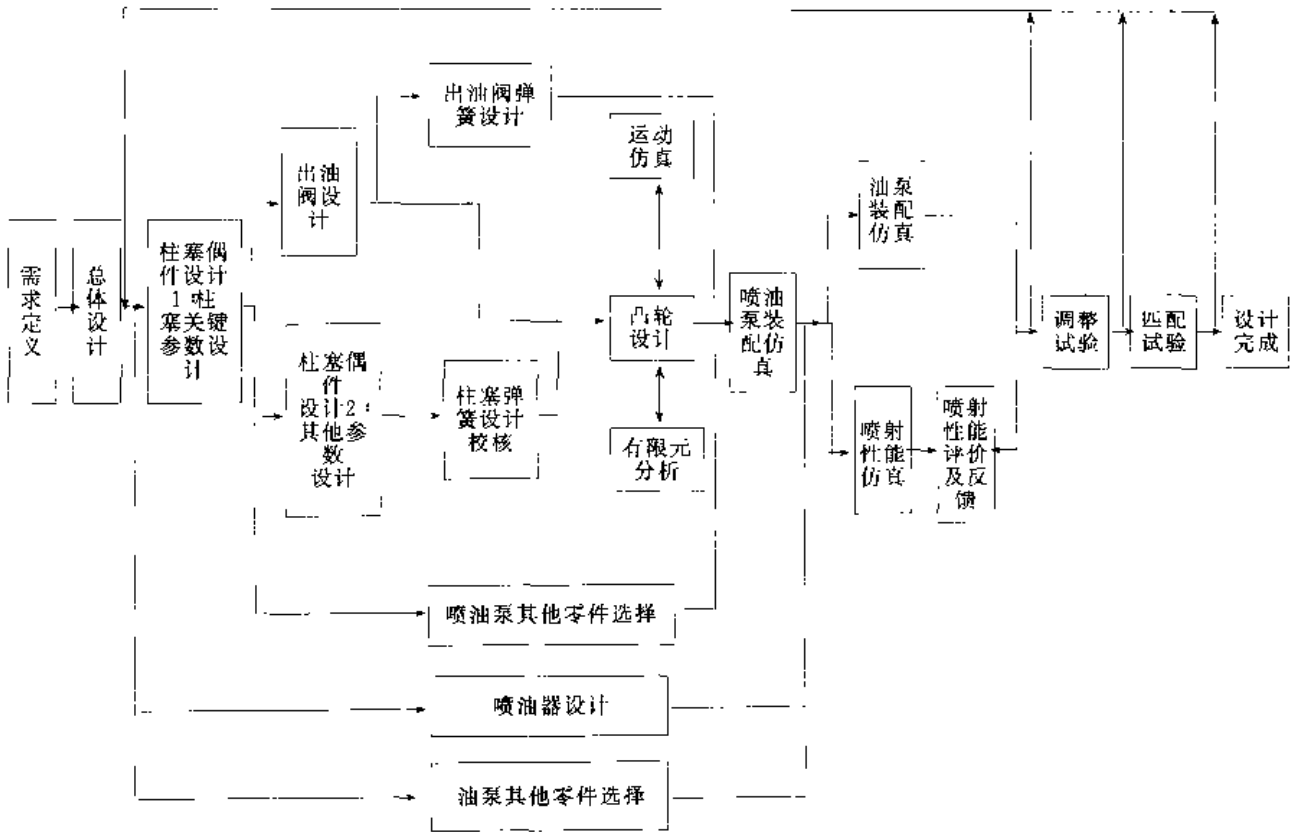


图 39.7-2 油泵并行开发过程

#### 1.4 电子仓库设计

在油泵类产品的研发过程中，特别是在三大偶件（柱塞偶件、出油阀偶件和凸轮偶件）及喷油器的设计阶段，产生了大量的设计数据，而且这些数据之间的偶合性非常强，因此更需要根据数据的实际传递情况对系统实施一个良好的 Vault 规划。

为了保证各设计组之间的数据既能按规定权限共享，又能保持一定的独立性，在系统实施中，将每个角色对应一个工作组，每个组拥有一个组内共享的电子仓库，这些电子仓库之间、以及他们与全局共享的电子仓库之间按照规定的路线传递数据。在设计过程中，三大偶件设计阶段及喷油器设计阶段需要的一些源数据来源于总体设计要求，因此它们对应的局域共享的电子仓库需要与总设计师组的电子仓库之间进行数据交换；而三大件自身的设计过程也不是相互独立的，出油阀偶件和凸轮偶件的设计依赖于柱塞偶件的一些设计参数（如柱塞直径等），因此三大偶件设计组对应的局域共享的电子仓库之间也存在着数据的传递。在柱塞偶件、出油阀偶件和凸轮偶件的设计完成后，比较成熟的设计参数传递给全局共享的电子仓库（“产品设计 Vault”）；仿真组和配试组的人员从“产品设计

Vault”中取得所需要的数据分别完成各自的工作；工艺人员则负责在设计参数被传递给“产品设计 Vault”之前对设计出的新型零部件进行可加工性检查，只有通过检查的设计方案才能进入全局共享库。在各项检查和试验都通过后，“产品设计 Vault”中的参数经组织后存档，放入“存档 Vault”中。总设计师组对应的电子仓库除了给三大偶件和喷油器的设计提供一些原始设计参数以外，还要和两个全局共享的电子仓库之间进行数据交互，以确保总师能够全面了解和监控设计过程的各个环节。各 Vault 具体分配及其相互之间的数据传递关系见图 39.7-3。其中箭头所指的方向代表了数据的传递方向，对应的操作可能是 View、All Transfer 或 Baseline 等。

#### 1.5 产品开发过程管理

产品开发过程管理是对开发过程中的活动 and 资源（人员、设备、应用工具等）进行实时监控和管理，并对开发过程中产生的冲突提供协调支持，以保证产品开发过程的顺利进行。该子系统包括三大部分。

##### (1) workflow 管理

workflow 管理是在 Metaphase 的生命周期管理器

(LCM) 的支持下进行的, 其主要功能包括:

1) 创建工作流(生命周期) 通过对规划好的产品开发过程进行分析, 将其演变成为由 LCM 基本作业模块(分配作业、评审作业、条件作业、消息作业、发布作业)构成的生命周期。

2) 执行工作流 将构成生命周期的作业模块变

成任务单, 分别指派给相关的人员, 由他们去完成, 执行完后返回生命周期, 由 LCM 指定执行下一任务。如果开发人员发现任务单元无法执行, 或在执行过程中出现问题, 就发出出错信息。

3) 历史管理 记录工作流的历史信息, 为信息查询和冲突协调提供参考。

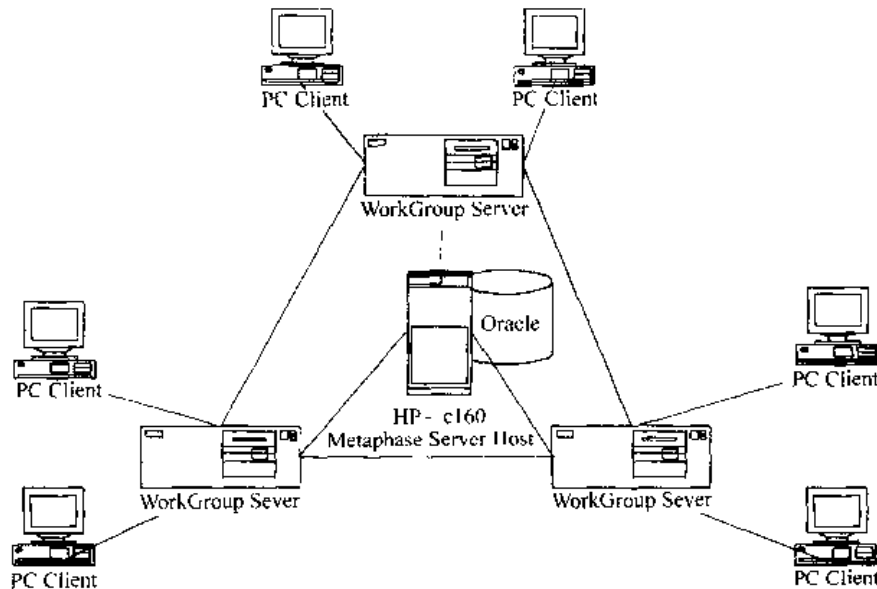


图 39.7-3 油泵并行设计拓扑结构

## (2) 过程监控

监视产品开发过程是否沿着要求的方向进行, 同时为开发人员提供评估和冲突等信息。主要功能有:

1) 进程监控 对产品开发过程各活动的进展情况实时监控, 同时还可结合各开发人员的个人工作进度评估, 预测整个过程的进展状况, 一旦发现某个活动出现问题, 就及时向管理人员提出警示信息。

2) 产品数据监控 监视所要求的产品数据是否产生, 是否存放于规定数据库中, 以及是否传递给了应得到数据的人。

3) 资源监控 资源监控包括用户需求资源监控、物力资源(设备、应用工具等)监控和人力资源监控三个方面。通过这些资源的监控和冲突的及时发现和解决, 为开发过程的顺利进行提供必要的支持。

4) 冲突检测 将进程、资源和产品数据的监控信息与所要求的信息进行比较, 判断是否存在冲突, 一旦检测到, 就提交冲突协调模块, 谋求解决方案。

## (3) 冲突协调

冲突可能产生于并行开发过程的各个阶段, 具有不同的表现形式。本模块的目的就是根据不同的冲突, 寻求不同的解决方案。其主要功能包括:

1) 冲突分析 根据冲突信息来源, 冲突表现形

式、冲突出现阶段等冲突属性, 判定冲突类别, 并由一定的规则选择相应的冲突协调方法;

2) 冲突协调 根据冲突类别的不同, 建立不同的协调策略和协调方法, 形成具有一定的自学习的冲突协调库, 以满足并行工程的需求。

## 1.7 基于 PDM 的油泵产品并行开发原型系统

### 1.7.1 油泵产品并行开发过程的活动流

通过 Metaphase 的 OMF (对象管理框架) 和 LCM (生命周期管理器), 将这些实例用 Metaphase 的模型表示出来。图 39.7-4 是油泵产品并行开发的活动流, 该活动流是由多个生命周期阶段 (Life Cycle, 以下简称 LC) 层层嵌套而得到的, 所以只有四个 LC。图 39.7-5 是该活动流所包含的一个 LC——柱塞弹簧设计 LC 的活动流, 它由三个进程 (Process) 构成。

图 39.7-6 表示的是油泵开发小组角色与开发活动的关系, 而图 39.7-7 则是成员与角色的关系。这些子视图都是利用 Metaphase 已有的信息模型实现的。

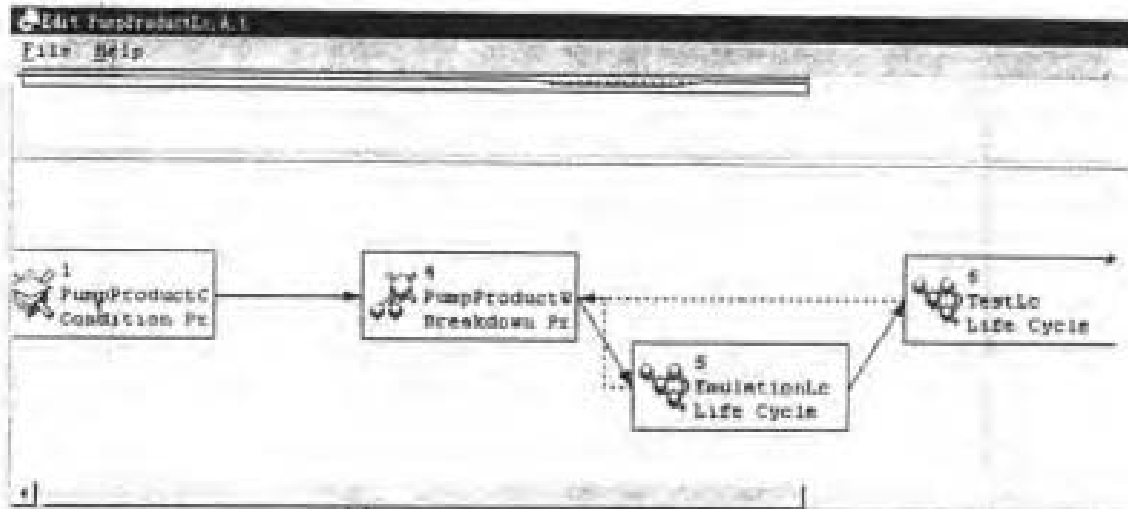


图 39.7-4 油泵产品并行开发过程的活动流

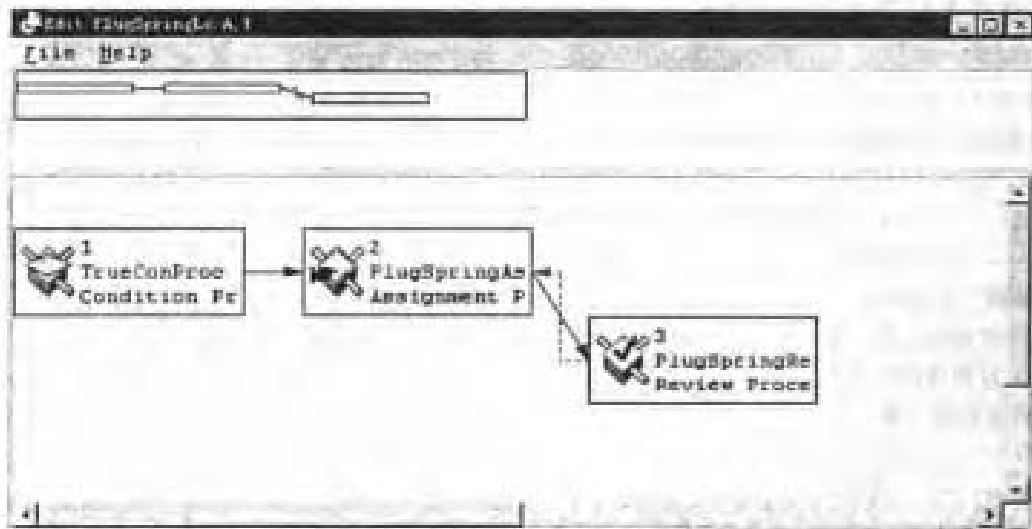


图 39.7-6 任意弹簧设计活动流

Item Identifier	Class	Role Description
CamDesigner	Role	---
Is Assignee In (acts)		
CamAssignProc, A, 1	Assignment Process	---
CamFemAnalysisAssignProc, A, 1	Assignment Process	---
EmulationEngr	Role	---
EmulationLeader	Role	---
Is Reviewer In (acts)		
EmulationRevProc, A, 1	Review Process	---
GeniDesigner	Role	---
HeadManager	Role	---
MuzzleDesigner	Role	---
MuzzleLeader	Role	---

图 39.7-6 IPDT 角色与开发活动的关系

Item Identifier	Class	Role Name	Project Name
Town	Role Assignment	TechnicianEngt	*
Town	Role Assignment	EvolutionLeader	*
Town	Role Assignment	HeadManager	*
Town	Role Assignment	CaseDesigner	*
Town	Role Assignment	GenDesigner	*
City	Role Assignment	RussiaLeader	*
City	Role Assignment	CaseDesigner	*
City	Role Assignment	RussiaDesigner	*
City	Role Assignment	ValveDesigner	*
City	Role Assignment	EvolutionEngt	*
Job	Role Assignment	TestEngt	*
Job	Role Assignment	PumpLeader	*
Job	Role Assignment	TestLeader	*

图 39-7-7 成员与角色的关系

### 1.7.2 产品开发过程监控器功能

产品开发过程监控器(Product Development Process Monitor, PDDM)是在 Metaphase 上二次开发的功能模块,实现了过程监控、冲突检测和冲突协调等功能,图 39-7-8 为其主界面,主菜单中包括 Requery (对象查询)、Info (对象一般信息)、Process (进程信息)、Lifecycle (生命周期信息)、Resource (资源信息,含成员信息)和 Conflict (冲突检测与协调)等功能项。

过程监控包括对监控对象静态信息的查询和对其动态信息的实时获取两个方面。

#### (1) 过程监控的对象

在 PDDM 中,监控对象的获取过程是:通过 PDDM→Requery→... 选取某类对象,再通过进一步约束对象的属性而获取所需对象。在 PDDM 中,被监控的对象包括:进程(分为所有进程、作业指派进程、评审进程等)、活动流(分为生命周期 LC、任务流)。

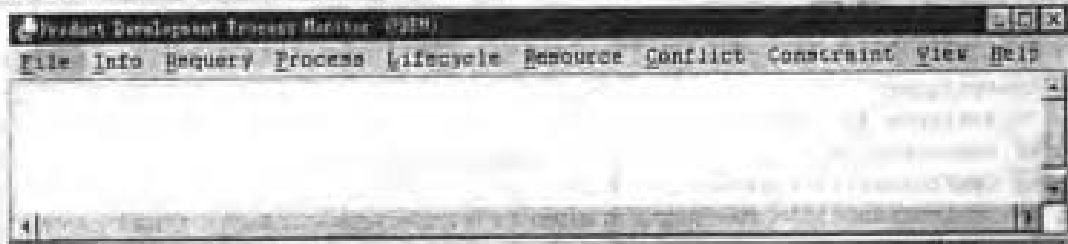


图 39-7-8 PDDM 的主界面

#### (4) 监控对象动态信息的实时获取

监控对象动态信息的获取方法与其静态信息的查询方法相同,可根据监控对象的类型在 PDDM 的相应菜单下得到。图 39-7-12 是某进程动态信息的输出界面,该动态信息描述了此进程当前所在的生命周期(因为一个进程可定义在多个生命周期中);是该生命周期的第几步,正在处理的数据 (Item)、进程的执行人

用户需求、用户以及约束网络等类型,产品数据对象在 PSM 模块中获取。图 39-7-9 为对应的菜单,界面还显示了已获取的所有进程对象。

#### (2) 监控对象信息的查询

Metaphase 对每一个管理对象都定义了相应的属性,这些属性的描述信息存储在 Info 类中,可通过查询 Info 菜单而得到,本文称对象的这类信息为一般信息,监控对象的一般信息也可通过 PDDM 的 Info 菜单而得到。

#### (3) 监控对象静态信息和历史信息的查询

根据监控对象的不同,其静态信息和历史信息也不相同,这就需要借助 Metaphase 的开发工具进行定义,查询时可根据监控对象的类型在 PDDM 的相应菜单下得到。图 39-7-10 是进程 (process) 的菜单,通过此菜单可得到进程的三类信息:静态信息、动态信息和历史信息。图 39-7-11 是某进程静态信息的输出界面(部分)。

员,工作进度等实时信息。

特别地,通过选择“Request latest rate of process”项,可直接向进程的执行人询问当前的工作进度。图 39-7-13 是进程执行人收到的要求他更新工作进度的消息,图 39-7-14 是工作进度的输入界面,输入得到确认后,该信息将立刻反映到动态信息中,通过刷新该进程的动态信息就可看到新的工作进度。

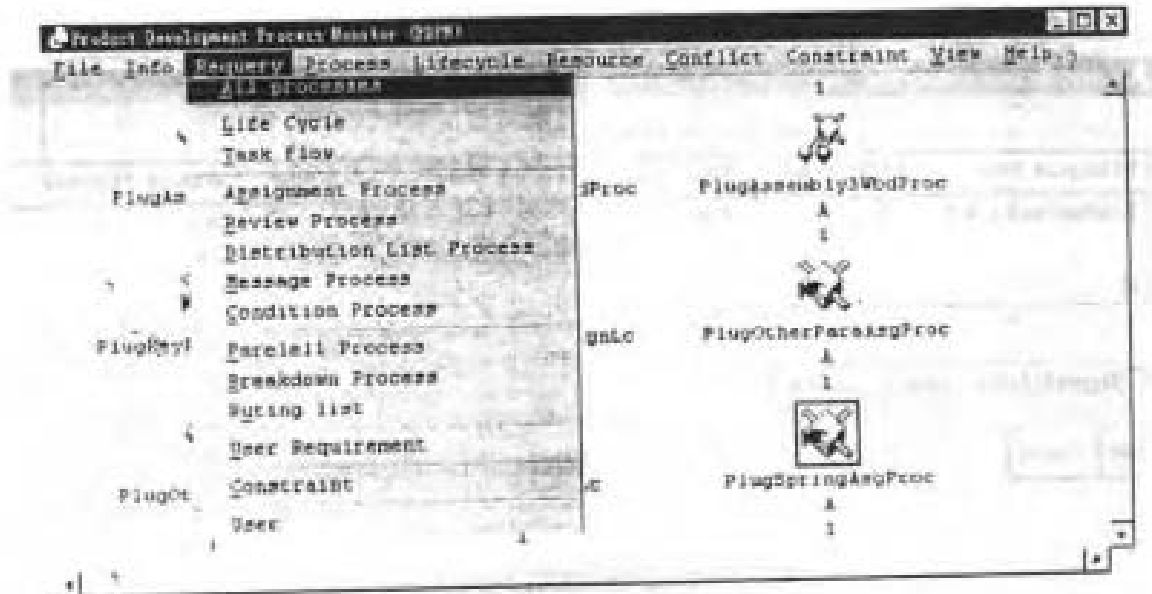


图 39.7-9 监控对象的获取菜单

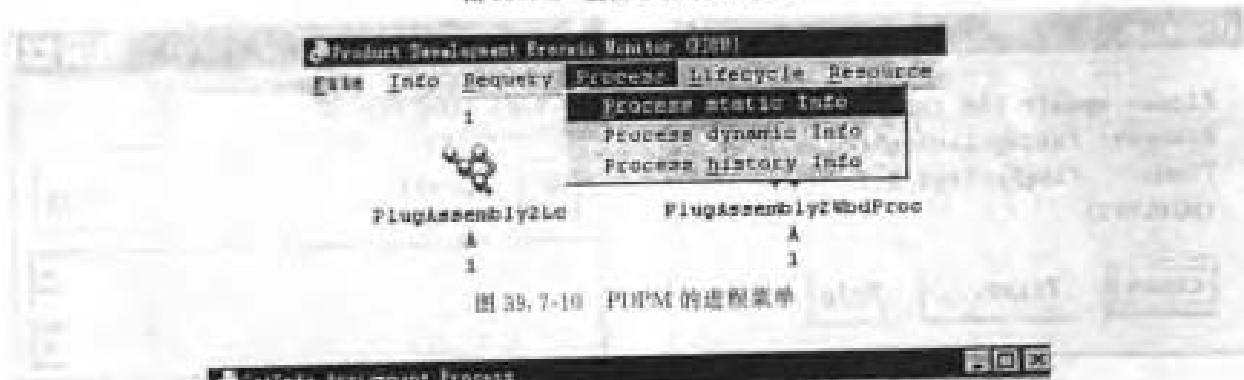


图 39.7-10 PDM 的进程菜单

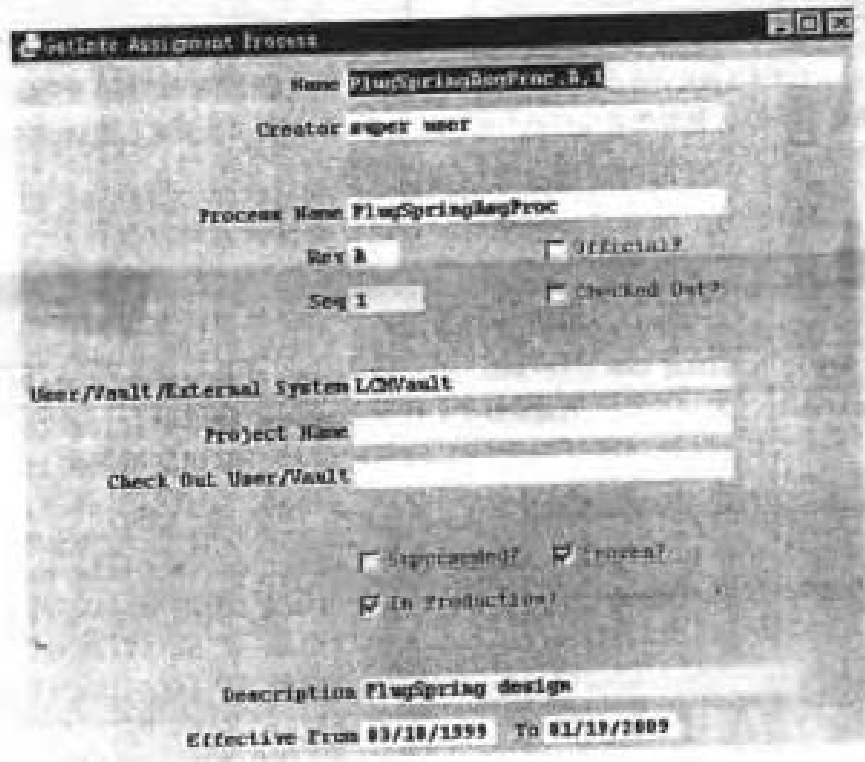


图 39.7-11 进程动态信息的输出界面 (部分)

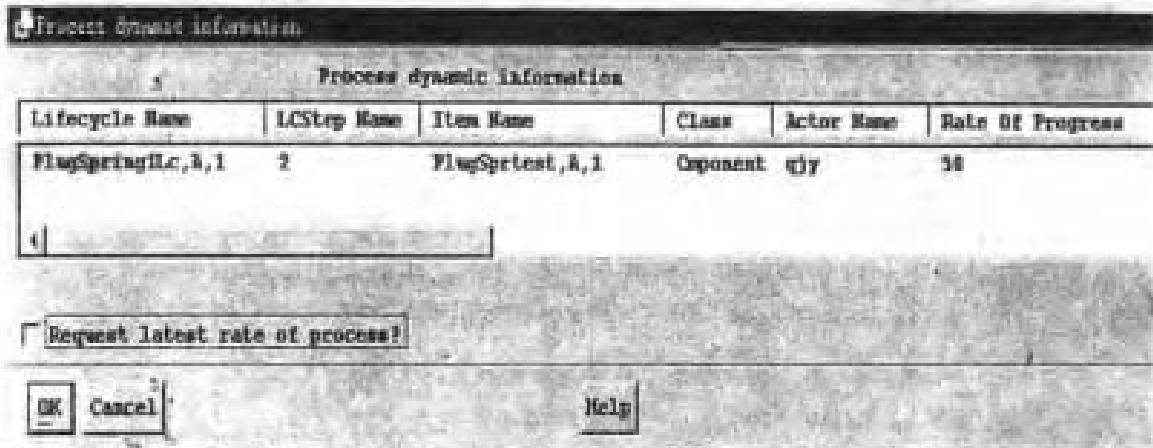


图 39-7-12 进程动态信息的输出界面

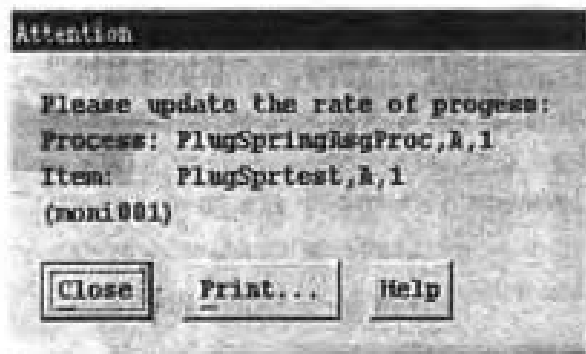


图 39-7-13 工作进度的更改通知

其他对象，如成员、活动流、资源的动态信息的获取与此类似，不再赘述。此外，PDPM 还可随时接受

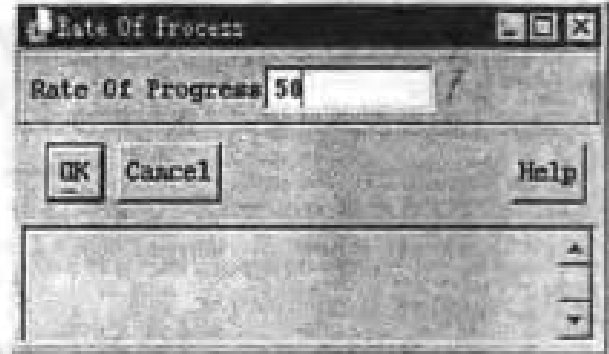


图 39-7-14 工作进度的输入界面

成员对资源需求的要求。图 39-7-16 是用户需求的输入界面。

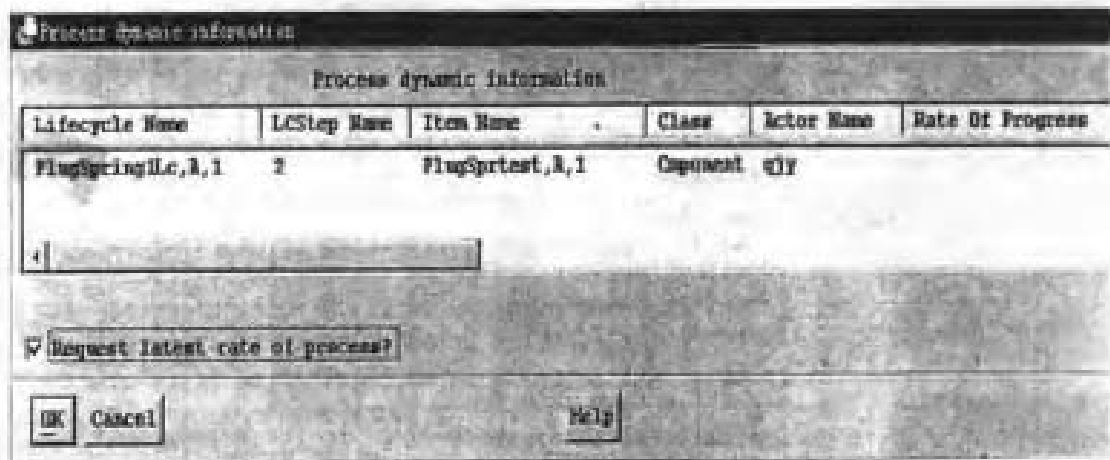


图 39-7-15 某进程刷新后的动态信息

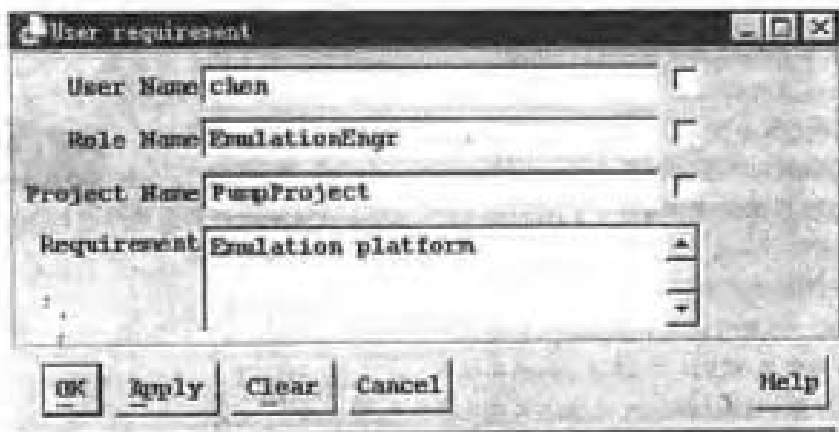


图 39-7-14 用户需求输入界面

## 2 基于 PDM 的客车车身并行设计

### 2.1 客车车身并行设计开发项目概况

#### (1) 项目目的和内容

在汽车行业,我国的企业与国外的先进企业相比,存在着较大的差距。在客车车身的设计方面,基本上还沿袭传统的串行设计方法和步骤,从绘制 1:5 客车车身布置图,绘制彩色效果图,雕塑 1:5 模型,到绘制 1:1 线型图,雕塑 1:1 油泥模型,制作 1:1 内部模型,绘制客车车身主图板,绘制客车车身零件图,样车

试制,样车试验,制造客车车身主模型,试验,加工生产等一系列串行活动及其大反馈在国内至少需要 5~6 年的时间才能完成,而且设计可靠性还很低,造成这种现象除了是由于设计流程的规划不合理以外,另一个重要原因就是企业内的信息流通不畅,设计、工艺和制造部门的人员不能共享产品信息,不能就各自环节中出现的问題及时交换意见,导致了許多无效的重复劳动以及开发活动中的大反馈,这样,无形中延长了开发周期,降低了客车车身的设计质量,所以很有必要进行并行设计的研究和实施工作。

随着产品市场的全球化,国外汽车不断涌入国内

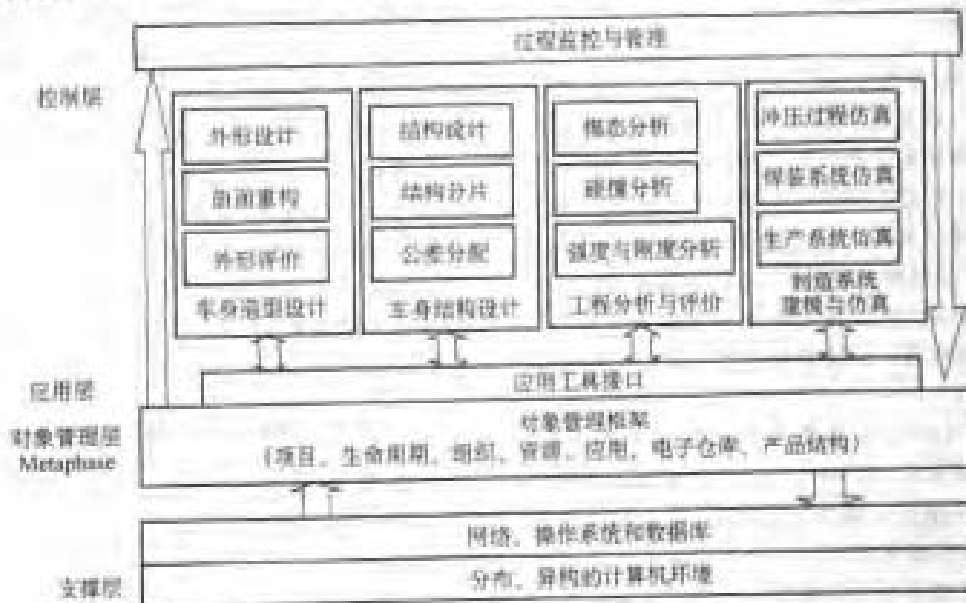


图 39-7-17 系统结构图

市场,国内汽车企业正面临着越来越激烈的冲击,为了适应市场变化,增强竞争力,国内汽车企业采用企业级 PDM 软件来支持并行设计设计已经成为必然趋势。

本项目研究开发的目的是:针对汽车客车车身设

计与开发全过程中概念设计、造型设计、结构设计、内饰设计和工艺设计等重大关键技术问题,建立并行设计的应用系统,并开展并行设计产品开发设计过程中知识的获取,分布异构环境下过程建模和集成等相关技术的攻关,设计出国内领先水平的新型客车的电子

样车。结合轿车客车车身改型设计和制造质量改进的实际需求,提高客车车身的设计质量和制造质量。

### (2) 原型系统的系统结构

面向创新设计的汽车客车车身并行设计开发系统由客车车身概念设计子系统、客车车身造型设计子系统、客车车身结构设计子系统、客车车身内饰设计子系统、客车车身工艺设计子系统和分布异构的集成系统组成。

原型系统的系统结构如图 39.7-17 所示。系统结构分为控制层、应用层、对象管理层和支撑层。控制层用来对产品开发的流程进行规划与控制,包括角色的定义、任务的分解与分配、流程的建立、反馈的控制等

等。应用层主要是与客车车身相关的各种 CAX 应用工具及其在系统中的封装,包括工具类定义,应用文件类定义,用户界面的定制等等。对象管理层是对与产品开发有关的各种对象进行管理,包括文件对象、产品结构对象、文档对象以及数据仓库对象等等。支撑层是用来支持系统运行的软硬件环境,包括网络、分布式数据库、操作系统等等。

## 2.2 PDM 系统结构

表 39.7-1 列出了 Metaphase 软件的安装节点的主机配置。

表 39.7-1 主机配置

机器名	机器类型	IP 地址	物理区域
VM_IRIS	SGI 图形工作站		C 研究所
Adams2	PC		C 研究所
Ansys3	PC		A 研究所
Chenwei	PC		A 研究所
Cims	PC		C 研究所
Icem	PC		A 研究所
Pdmpc1	PC		C 研究所
Pdmpc2	PC		C 研究所
Pdmserver	PC		C 研究所
Rp2	PC		B 研究所
Vr	PC		C 研究所
SIMPLE			C 研究所
KEREN SERVER			B 研究所

## 2.3 逻辑拓扑

由于在 PDM 系统中采用 TCP/IP 通信协议,因此要求系统的每一台主机都有唯一的 IP 地址,并且能够通过 TCP/IP 相互连接到对方。组成系统的主机虽然都连接在物理上连通的网络上,但是由于彼此分别隶属于不同的内部局域网,并且有防火墙将其与外部网隔离,造成一个局域网内部的 IP 地址不能被另一个局域网中的主机识别(如 C 研究所和 A 研究所之间的关系),形成通信障碍,因此需要在系统中建立特定的 PDM 网络视图。

在由 PDM server 及传统客户端组成的系统中,共分成 3 个网络视图(图 39.7-18)。划分网络视图的原则是凡是相互识别并连通对方 IP 地址的机器可以放在同一个网络视图中,一个大视图也可以根据地城关系再细分为若干个子视图;凡是不能相互识别并连通对方 IP 地址的机器必须放在不同的网络视图中,网络视图之间的通信通过网桥实现。

## 2.4 文件系统定义

文件系统是 Metaphase 软件用来存放应用数据的地方,一个文件系统对应操作系统中某一个指定的目

录及其下属的所有子目录。只有处于文件系统对应目录下的文件才能被 PDM 系统管理。一般来说,用户的文件在共享之前都放在个人工作区,个人工作区必须放在文件系统对应的目录或者其子目录中,因此为了便于管理用户的个人数据,必须在每台主机上都至少设定一个文件系统。但是文件系统太多的话,也会给管理带来不便,因此在本项目对应的支撑系统中,除个别用于测试或特殊用途的主机外,每台主机上设置且仅设置一个文件系统。图 39.7-19 列出了部分文件系统及其与主机的对应关系。

## 2.5 Client/Server 与 Web/Browser 结构

支撑平台采用 Client/Server 结构及 Web/Browser 结构。在主机群中,主机 pdmserver 是整个系统的核心。

一方面,pdmserver 是 PDM 系统的服务器,所有 Metaphase 的核心服务都在该主机上运行,客户端只需运行与操作系统有关的 OS Server 及与图形用户界面相关的 GUI Server。在服务器端运行正常、客户机与服务器之间连接正确的情况下,每个客户端在启动的时候和服务器端建立连接,从服务器端读取系统配置信息;各主机之间通过专用的 Metaphase 通信进程



及任务调度进程 Dispatcher 进行交互,数据库系统也安装在 pdmserver 上。

另一方面, pdmserver 也是支撑系统的 Web 服务器,内部运行着 WWW 服务,并且在 WWW 服务与 PDM 服务之间建立了紧密的联系。

最后, pdmserver 还是承担网关及小组级服务器的功能。C 所局域网内部的机器与所有外部机器的通信通过 pdmserver 中介完成,局域网外部的主机也只有通过 pdmserver 才能与 C 所内的主机进行通信,除了核心服务器以外,还有小组级 PDM 服务器,这里将

配置有 Vault 的主机成为小组级服务器,全局共享 Vault 级多级小组级 Vault 均放置于主机 pdmserver 上,其余 Vault 放置在主机 icem 上。

### 2.6 用户、角色定义

系统中共定义了 13 个用户,11 个角色。见图 39.7-20。

用户与角色、用户组,主机之间的对应关系见表 39.7-2。

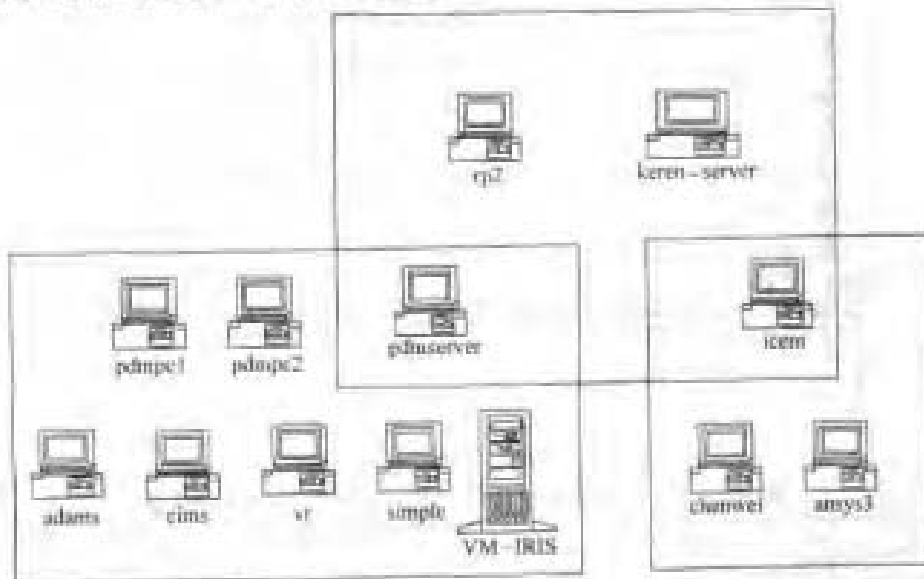


图 39.7-18 网络视图及相互间关系

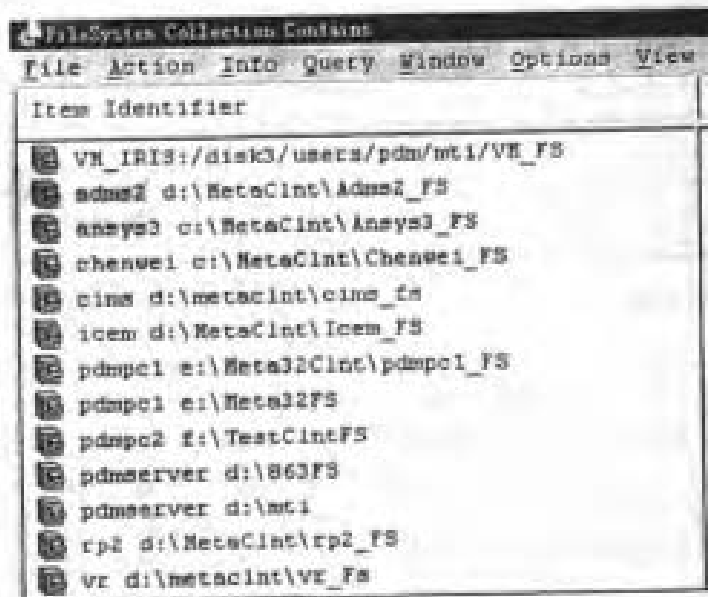


图 39.7-19 文件系统

Users Collection Contents		Roles Collection Contents	
File	Action	Info	Query
Item Identifier		Item Identifier	
■	BaoJS	■	ChongBanEngineer
■	HuangB	■	DefaultChangeAdmin
■	JiangZH	■	DefaultChangeAnalyst
■	JinXL	■	DefaultReviewBoard
■	LaiXM	■	FenXiDesigner
■	LiYX	■	GongWeiEngineer
■	WangHL	■	JieGouDesigner
■	WangW	■	NeiShiDesigner
■	WangYL	■	VR_Simulator
■	XiJT	■	WaiXingDesigner
■	XuWL	■	ZongBuZhiDesginer
■	lyj		
■	super user		
13 objects found.		11 objects found.	

图 39.7-20 用户、角色定义

表 39.7-2 用户与角色、用户组、主机之间的对应关系

用户	角色	可登录主机	所在用户组
BaoJS	VR-Simulator	Vr, VM_IRIS	user_grp, 863Group
HuangB	NeiShiDesigner	Cims, adams2	user_grp, 863Group
JiangZH	NeiShiDesigner	Cims, adams2	user_grp, 863Group
JinXL	ZongBuZhiDesigner	icem	user_grp, 863Group
LaiXM	ChongBanEngineer	icem	user_grp, 863Group
LiYX	WaiXingDesigner	Ansys1	user_grp, 863Group
WangHL	JieGouDesigner, FenXiDesigner	icem, chunwei	user_grp, 863Group
WangW	GongWeiEngineer	SIMPLE	user_grp, 863Group
WangYL	ZongBuZhiDesigner	KEREN, SERVER	user_grp, 863Group
XiJT	NeiShiDesigner	Rpl	user_grp, 863Group
XuWL	ChongBanEngineer	icem	user_grp, 863Group
LiYJ	System Administrator	Pdmserver, pdmipc1, pdmipc2	Super user grp
Super user	System Administrator	Any host	Super user grp

## 2.7 Vault 结构及权限控制

### (1) 个人工作区 WorkLocation

在 Metaphase 软件中，个人工作区称为 Work Location。一般情况下，每个设计、开发人员在他能够登录的主机上至少拥有一个个人工作区，用于存放私有数据。由于个人工作区是用来管理每个人自身的私有数据的，所以 work location 的管理权完全由各个用户自己控制。用户可以根据自己的需要创建一个或者多个 work location，在存放一个数据文件时指定个人工作区作为存放位置。

由于个人工作区的灵活性及多样性，这里不对其进行列表。图 39.7-21 给出了部分 Work Location 及其所有者示例。

### (2) Vault Location 及 Vault 列表

Vault 是用来存放共享信息的地方，对于元数据 (MetaData, 用来描述数据的数据) 来说，其信息存放于 Oracle 数据库中，不需与文件系统发生关系；对于应用数据来说，其信息存放于文件系统中，因此需要指定具体的存放子目录。在 Metaphase 中，这一概念称为 Vault Location。与一个用户可以拥有多个个人工作区 Worklocation 一样，一个 Vault 可以拥有一个或多

个 Vault Location。

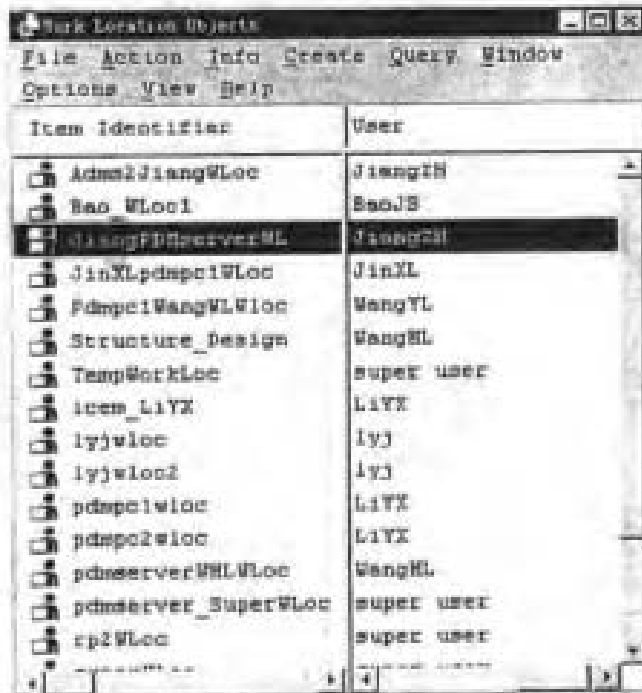
系统中涉及的 Vault Location 及其与 Vault 之间的关系见图 39-7-22。

### (3) 权限控制

为了实现应用数据的合理共享,采用如图 39-7-23 所示的多级权限规划。

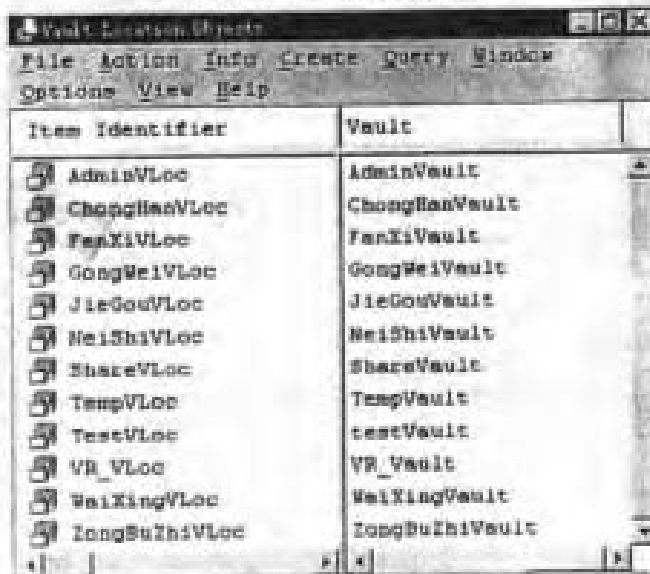
其中,个人工作区的数据由用户个人全权控制,局部共享区内应用数据由小组成员集体控制,即小组内的每一个成员都有权将自己个人工作区内的数据传送到自己所在的小组共享区,也可以将应用数据从小组共享区中检出 (CheckOut) 予以修改或者更新,但小

组成员无权将应用数据数据传递到其他小组对应的局部共享区,也无权将其他共享区内的应用数据检出 (CheckOut) 予以修改或者更新。全局共享区内的 AdminVault 用来对文档进行归档处理,由系统管理员进行数据传递和归档操作,或者在流程管理中通过专门的 Process 予以自动完成。为了便于小组成员之间在流程管理系统之外随时进行数据的必要传输,在全局共享区还设定了另外一个 Vault——ShareVault,所有的用户对这一 Vault 都有足够的权限,如数据检入 (Check In), 检出 (Check Out), 传递 (TransferIn and TransferOut), 删除 (Delete)。



Item Identifier	User
AdminJiangWLoc	JiangJH
Bao_WLoc1	BaoJB
JiangPDMserverWLoc	JiangJH
JinXLpdmpciWLoc	JinXL
PdmpciWangWLoc	WangYL
Structure_Design	WangHL
TempWorkLoc	super user
icem_LiYX	LiYX
lyjwloc	lyj
lyjwloc2	lyj
pdmpciwloc	LiYX
pdmpc2wloc	LiYX
pdmserverWHLWLoc	WangHL
pdmserver_SuperWLoc	super user
rp2WLoc	super user

图 39-7-21 Work Location 示例



Item Identifier	Vault
AdminVLoc	AdminVault
ChongBanVLoc	ChongBanVault
FanXiVLoc	FanXiVault
GongWeiVLoc	GongWeiVault
JieGouVLoc	JieGouVault
HeiShiVLoc	HeiShiVault
ShareVLoc	ShareVault
TempVLoc	TempVault
TestVLoc	testVault
VR_VLoc	VR_Vault
WeiXingVLoc	WeiXingVault
ZongBuZhiVLoc	ZongBuZhiVault

图 39-7-22 Vault Location 与 Vault

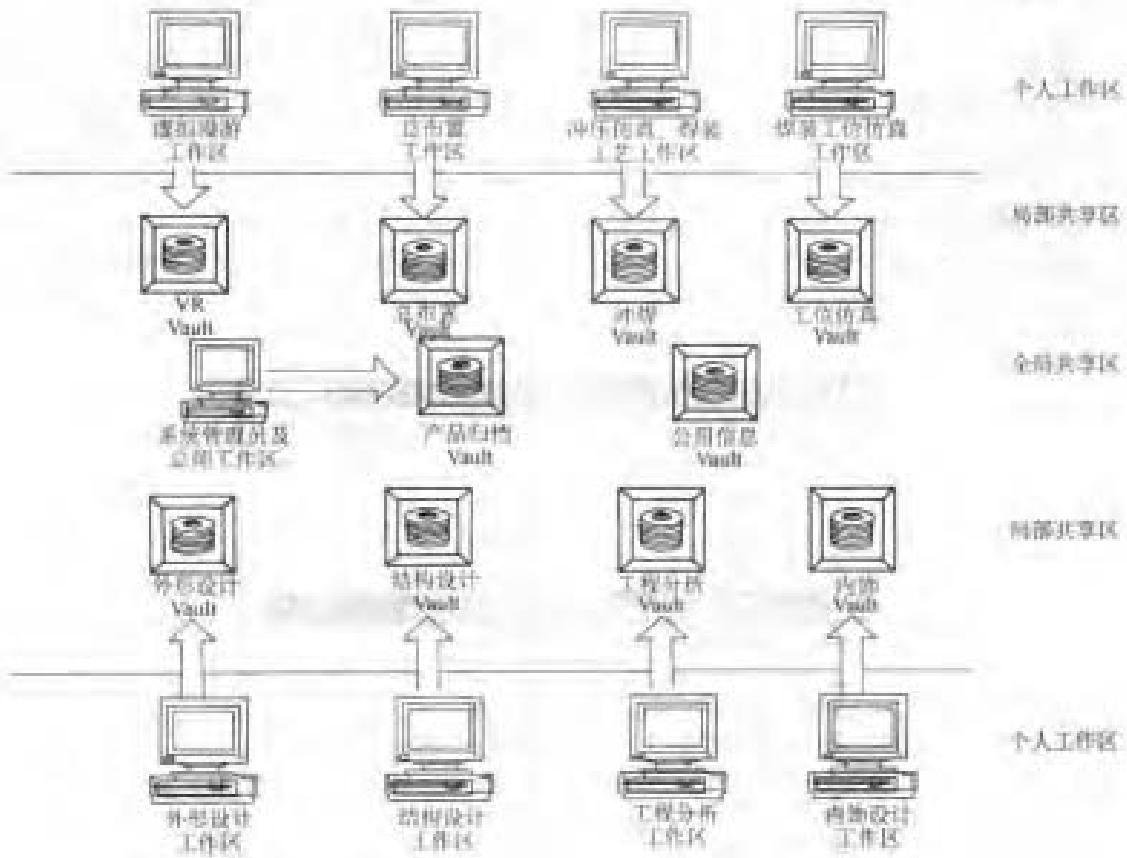


图 39-7-23 多用户权限设计

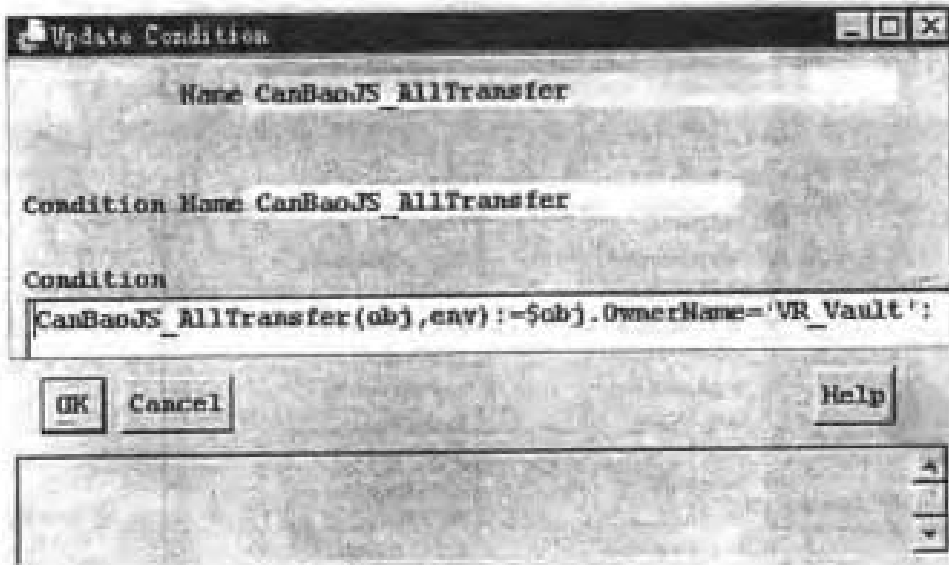


图 39-7-24 Condition 示例

为了便于管理,凡是 User group 内的用户,都赋予其浏览和查看文档的权限。对于个人工作区而言,如前所述,其权限完全由各个用户自己控制;对于 Vault 而言,其权限的控制是通过写条件(Condition)和消息可获取规则(Message Access Rules)实现的。

Condition 是用来决定某一事件或者操作是否能够进行的条件,Condition 本身并不能发挥作用,它需要与其他规则结合起来使用才有意义。在一个 PDM 系统中,通常需要书写大量的 Condition。Condition 所包含的内容实际是对某一对象的某些属性进行限定,图 39.7-24 给出了 Condition “CanBaoJS-AllTransfer”的具体格式及内容。在本系统中,与 Vault 权限控制有关的 Condition 名称列表见图 39.7-25。

#### (4) Message Access Rules 列表

在基于面向对象思想设计的系统中,一个操作对应一个消息,通过设定用户对消息的可获取性就可以控制用户对某一操作的可执行性。创建一个 Message Access Rules 的过程实际上就是给定几个参数,设定“某用户(用户组)在某对象满足某条件的情况下能够对其进行某种类型的操作”。

图 39.7-26 为 Message Access Rule 的创建界面。其中 Condition 一栏填写预先定义的条件名称,Class 一栏填写对象名称,Message Name/Message Group

Name 一栏填写消息名称(对应一个操作),Participant 一栏填写用户或者用户组名称。

以规则“All-transfers-grp, BaoJS”为例,它所决定的权限可以描述为:“用户在 WorkItem 及其子类满足条件‘CanBaoJS-AllTransfer’时,可以对其进行‘All transfers-grp’消息组所对应的一组操作。”该规则的书写格式及内容见图 39.7-27。系统中建立的 Message Access Rules 列表见图 39.7-28。

## 2.8 流程管理

### 流程分析与实现

#### (1) 流程文档设计(图 39.7-29)

#### (2) 过程及流程设计

##### 1) Condition Process 定义见表 39.7-3

##### 2) Assignment Process 定义见表 39.7-4

##### 3) Review Process 定义见表 39.7-5

##### 4) Vault Transfer Process 定义见表 39.7-6

##### 5) Taskflow 定义见表 39.7-7

##### 6) Parallel Process 定义见表 39.7-8

##### 7) Life Cycle 定义见表 39.7-9

##### 8) Breakdown Process 定义见表 39.7-10

#### (3) 可视化表达

##### 1) 一级流程,如图 39.7-30~图 39.7-32 所示。

表 39.7-3 设计文档管理规则

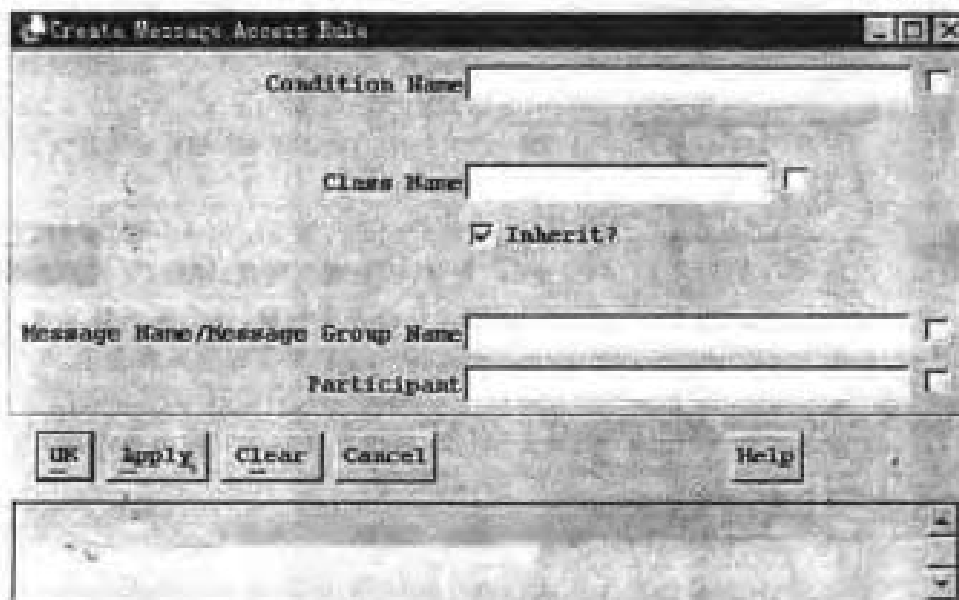
名称	说明	条件	备注
EntryPointCP	流程入口	TRUE	
Is_DeChair_Doc	只允许与驾驶员座椅相关的文档通过	Is_DeChair_Doc (itemobj, relobj, brkdwnobj); = \$ brkdwnobj. DocumentName='DriverChairDesDoc';	
Is_Deco_Doc	只允许与内饰设计相关的文档通过	Is_Deco_Doc (itemobj, relobj, brkdwnobj); = \$ brkdwnobj. DocumentName='DecoDesDoc';	
Is_Layout_Doc	只允许与总布局相关的文档通过	Is_Layout_Doc (itemobj, relobj, brkdwnobj); = \$ brkdwnobj. DocumentName='LayoutSpecDoc';	
Is_Outline_Doc	只允许与外形设计相关的文档通过	Is_Outline_Doc (itemobj, relobj, brkdwnobj); = \$ brkdwnobj. DocumentName='OutlineDesDoc';	
Is_Structure_Doc	只允许与结构设计相关的文档通过	Is_Structure_Doc (itemobj, relobj, brkdwnobj); = \$ brkdwnobj. DocumentName='StructureDesDoc';	
Return ToAccessoryDesignAP	多反馈中是否返回到附件设计	Return To AccessoryDesign AP (o, e); = \$ o. Which Process='AccessoryDesignAP';	
Return ToAerodynamicsSimu	多反馈中是否返回到空气动力学仿真	Return To AerodynamicsSimu (o, e); = \$ o. Which Process='AerodynamicsSimulationAP';	
Return ToOutLineAP	多反馈中是否返回到外形设计	Return To OutLineAP (o, e); = \$ o. Which Process='OutlineDesignAP';	



Item Identifier	Class	Creation Date
CanBaoJS_AllTransfer	Condition	07/19/2000 14:
CanHuangB_AllTransfe	Condition	07/19/2000 14:
CanJiangZH_AllTransf	Condition	07/19/2000 14:
CanJinXL_AllTransfer	Condition	07/19/2000 14:
CanLaiXM_AllTransfer	Condition	07/19/2000 14:
CanLiYX_AllTransfer	Condition	07/19/2000 14:
CanWangHL_AllTransfe	Condition	07/19/2000 14:
CanWangYL_AllTransfe	Condition	07/19/2000 14:
CanXiJT_AllTransfer	Condition	07/19/2000 14:
CanXuWL_AllTransfer	Condition	07/19/2000 14:
IsInAdminVault	Condition	07/19/2000 9:4
IsInChongHanVault	Condition	07/19/2000 9:4
IsInFenXiVault	Condition	07/19/2000 9:4
IsInGongWeiVault	Condition	07/19/2000 9:5
IsInJieGouVault	Condition	07/19/2000 9:5
IsInWeiShiVault	Condition	07/19/2000 9:5
IsInShareVault	Condition	07/13/2000 10:
IsInVR_Vault	Condition	07/19/2000 9:5
IsInWaiXingVault	Condition	07/19/2000 9:5
IsInZongBuZhiVault	Condition	07/19/2000 9:5

20 objects found.

图 39.7-25 与 Vault 权限控制有关的 Condition 列表



Condition Name: \_\_\_\_\_

Class Name: \_\_\_\_\_

Inherit?

Message Name/Message Group Name: \_\_\_\_\_

Participant: \_\_\_\_\_

OK apply Clear Cancel Help

图 39.7-26 规则创建界面

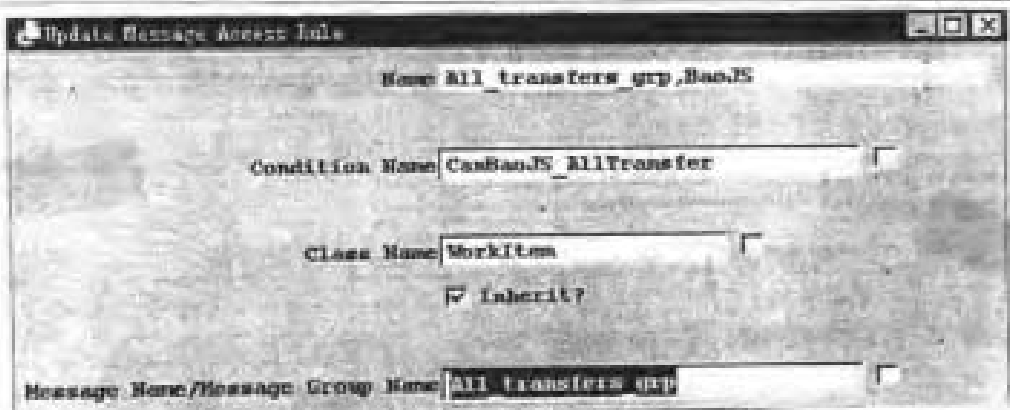


图 19.7-27 消息获取规则示例

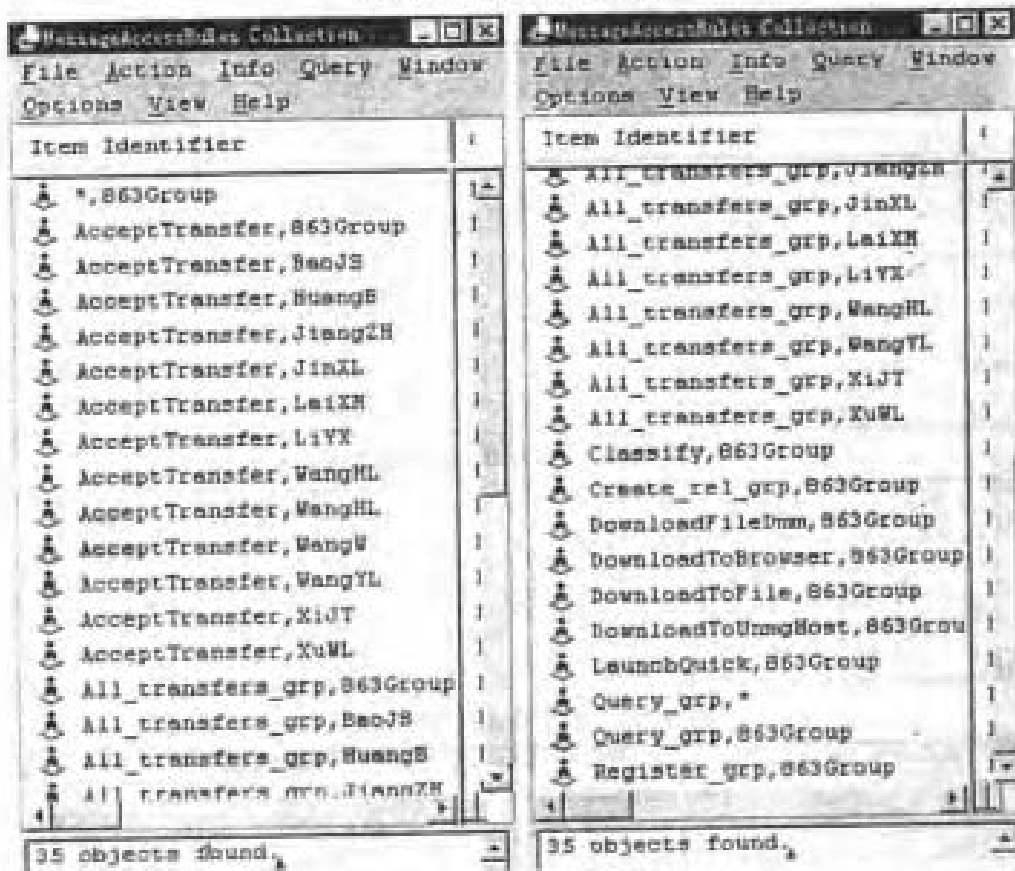


图 19.7-28 Message Access Rules 列表

Item Identifier	Class
863CarGenDoc	General Document
Depends On (dep)	
DecoDesDoc	Design Document
DriverChairDesDoc	Design Document
LayoutSpecfDoc	Specification Document
OutlineDesDoc	Design Document
StructureDesDoc	Design Document
VRDesDoc	Design Document

图 30.7-99 并行设计文档的存储管理

表 39.7-4 设计任务分配过程定义

名称	说明	执行者	备注
SelectStyleAP	客车车身造型	WangYL	
ReviseStyleAP	客车车身改型	JinXL	
DynamicSimulationAP	整车动力性能仿真		
FuelConsumeSimulationAP	整车燃油性能仿真		
BrakeSimulationAP	整车制动性能仿真		
SmoothSimulationAP	整车平顺性能仿真		
OperationSimulationAP	整车操作稳定性仿真		
Passability SimulationAP	整车通过性能仿真		
OutlineDesignAP	客车车身后外观设计	LiYX	
AerodynamicsSimulationAP	空气动力学仿真		
AccessoryDesignAP	客车车身附件设计	LiYX	
SkeletonDesignAP	客车车身骨架设计	WangHL	
StaticAnalysisAP	客车车身骨架静态特性分析	WangHL	
DynamicAnalysisAP	客车车身骨架动态特性分析	WangHL	
BumperAnalysisAP	客车车身骨架碰撞分析	WangHL	
PunchSimulationAP	冲压仿真	XuWL	
WeldProSimulationAP	焊装工艺仿真	LaiXM	
WeldPosSimulationAP	焊装工位仿真	WangW	
ProdLineSimulationAP	生产线仿真	WeiD	
DriverChairSurfREAP	驾驶员座椅曲面反求	WangJJ	
DriverDesignAP	驾驶员座椅造型	HuangB	
DriverChairAnalysisAP	驾驶员座椅平顺性仿真	HuangB	
DecoDesignAP	内饰造型与变形设计	JiangZH	包括乘客座椅、仪表盘、操纵杆、方向盘等
DecoSimulationAP	内饰仿真	BaoJS	包括视野、可及性、舒适性等
VirtualReality AP	并行设计漫游建模	BaoJS	

表 39.7-5 设计活动评审过程定义

名称	说明	执行者	备注
PerformanceRP	整车动力性能评价	JinXL	
OutlineRP	客车车身后外观设计评价	JinXL	
SkeletonRP	客车车身骨架设计评价	JinXL	
DriverChairSurfRP	驾驶员座椅曲面评价	WangJJ	
DecorationRP	内饰评价与模糊再设计	JiangZH	
VirtualRealityRP	并行设计漫游评价	BaoJS	

表 39.7-6 数据仓库转换过程定义

名称	消息	目标 Vault	备注
TransferToZongBuZu	LcmVaultTransItemObj	ZongBuZhiVault	
TransferToAdmin	LcmVaultTransItemObj	AdminVault	
TransferToWaiXing	LcmVaultTransItemObj	WaiXingVault	
TransferToJieGou	LcmVaultTransItemObj	JieGouVault	
TransferToNeiShi	LcmVaultTransItemObj	NeiShiVault	
TransferToVR	LcmVaultTransItemObj	VR_Vault	

表 39.7-7 任务流定义

名称	说明	选择条件	文件类型	包含的 Processes
DynamicsAnalysisTf	整车动力性能分析	TRUE	BI	EntryPointCP DynamicSimulationAP
FuelConsumeAnalysisTf	整车燃油性能分析	TRUE	BI	EntryPointCP FuelConsumeSimulationAP
BrakeAnalysisTf	整车制动性能分析	TRUE	BI	EntryPointCP BrakeSimulationAP
SmoothAnalysisTf	整车平顺性能分析	TRUE	BI	EntryPointCP SmoothSimulationAP
OperationAnalysisTf	整车操作稳定性分析	TRUE	BI	EntryPointCP OperationSimulationAP



(续)

名称	说明	选择条件	文件类型	包含的 Processes
PassabilityAnalysisTf	整车通过性能分析	TRUE	BI	EntryPointCP PassabilitySimulationAP
StaticAnalysisTf	结构静态特性分析	TRUE	BI	EntryPointCP StaticAnalysisAP
DynamicAnalysisTf	结构动态特性分析	TRUE	BI	EntryPointCP DynamicAnalysisAP
BumperAnalysisTf	结构碰撞分析	TRUE	BI	EntryPointCP BumperAnalysisAP

表 39.7-8 并行工作的流程定义

名称	说明	选择条件	包含的 Taskflow
PerformanceSimulationPP	整车性能仿真	TRUE	DynamicsAnalysisTf
		TRUE	FuelConsumeAnalysisTf
		TRUE	BrakeAnalysisTf
		TRUE	SmoothAnalysisTf
		TRUE	OperationAnalysisTf
		TRUE	PassabilityAnalysisTf
StructureAnalysisPP	客车车身结构性能分析	TRUE	StaticAnalysisTf
		TRUE	DynamicAnalysisTf
		TRUE	BumperAnalysisTf

表 39.7-9 生命周期阶段定义

名称	说明	包含的 Processes			
		Node	Process	Success	Fail
VirtualCarDesignLC	并行设计客车车身开发	1	EntryPointCP	2	Exit
		2	LayoutDesignBP	3	Exit
		3	OutlineDesignBP	4	Exit
		4	DecorationDesignBP	5	Exit
LayoutDesignLC	客车车身总布置设计	1	EntryPointCP	9	Exit
		9	TransferToZongBuZhi	2	Exit
		2	SelectStyleAP	3	Exit
		3	ReviseStyleAP	4	Exit
		4	PerfomanceSimulationpp	5	2
OutlineDesignLC	客车车身外形设计	5	TransferToAdmin	Exit	
		1	EntryPointCP	8	Exit
		8	TransferToWaiXing	2	Exit
		2	OutlineDesignAP	3	Exit
		3	AerodynamicsSimulationAP	4	Exit
		1	AccessoryDesignAP	5	Exit
		5	OutlineRP	6	10
		6	TransferToAdmin	Exit	Exit
		10	ConProAccessoryDesignAP	4	11
		11	ConProAerodynamicsSimu	3	12
StructureDesignLC	客车车身结构设计	12	ConProOutLineAP	2	Exit
		1	EntryPointCP	5	Exit
		5	TransferToJieGou	2	Exit
		2	SkeletonDesignAP	3	Exit
		3	StructureAnalysisPP	4	Exit
		4	SkeletonRP	6	2
ReverseEngLC	逆向工程	6	TransferToAdmin	Exit	Exit
		1	EntryPointCP	2	Exit
		2	DriverChairSurfCEAP	3	Exit
		3	DriverChairSurfRP	Exit	2

(续)

名称	说明	包含的 Processes			
		Node	Process	Success	Fail
DiverChairDesignLC	驾驶员座椅设计	1	EntryPointCP	5	Exit
		5	TransferToNeiShi	2	
		2	ReverseEngLC	3	Exit
		3	DriverDesignAP	4	2
		1	DriverChairAnalysisAP	6	3
		6	TransferToAdmin	Exit	Exit
DecoDesignLC	内饰设计	1	EntryPointCP	5	Exit
		5	TransferToNeiShi	2	Exit
		2	DecoDesignAP	3	Exit
		3	DecoSimulationAP	4	Exit
		4	DecorationRP	6	2
		6	TransferToNeiShi	Exit	Exit
MfgAnalysisLC	可制造性分析	1	EntryPointCP	2	Exit
		2	PunchSimulationAP	3	Exit
		3	WeldProSimulationAP	4	Exit
		1	WeldPosSimulationAP	5	Exit
		5	ProdLineSimulationAP	Exit	Exit
VirtualRealityLC	并行设计漫游	1	EntryPointCP	1	Exit
		4	TransferToVR	2	Exit
		2	VirtualRealityAP	3	Exit
		3	VirtualRealityRP	5	2
		5	TransferToAdmin	Exit	Exit

表 39.7-10 活动中止的流程定义

名称	说明	包含的 LC	优先级	条件	备注
LayoutDesignBP	总布置设计	LayoutDesignLC	1	Is-Layout-Doc	
OutlineDesignBP	外形设计	OutlineDesignLC	1	Is-Outline-Doc	
StructureDesignBP	结构设计	StructureDesignLC	1	Is-Structure-Doc	
TempBP	并行设计节点	MfgAnalysisLC	1	FALSE	
		DiverChairDesignLC	2	Is-Dchair-Doc	
		DecoDesignLC	3	Is-Pchair-Doc	
VirtualRealityBP	并行设计漫游	VirtualRealityLC	1	Is-VR-Doc	

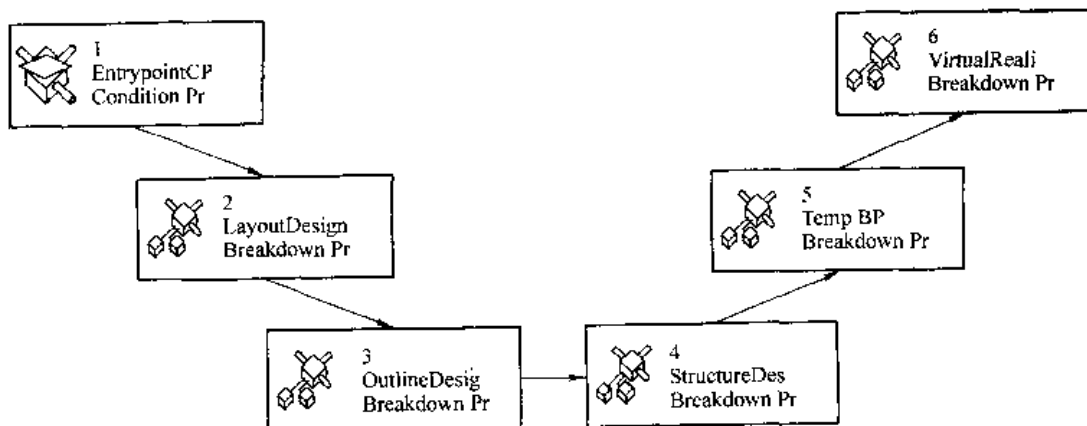


图 39.7-30 客车车身并行设计开发流程

2) 二级流程,如图 39.7-33~图 39.7-43 所示。

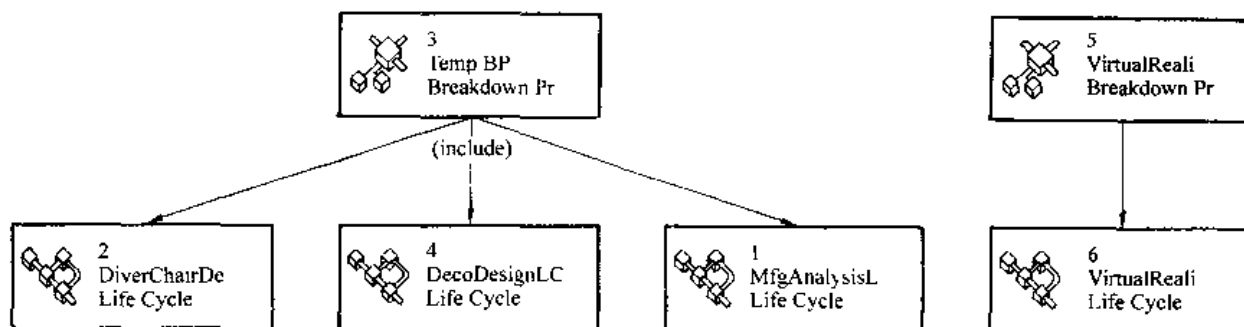


图 39.7-31 客车内饰布置设计开发流程

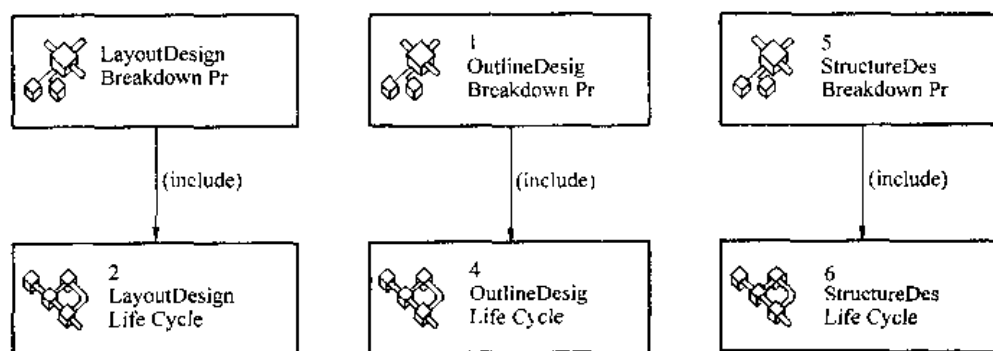


图 39.7-32 设计流程中止与设计活动的对应性

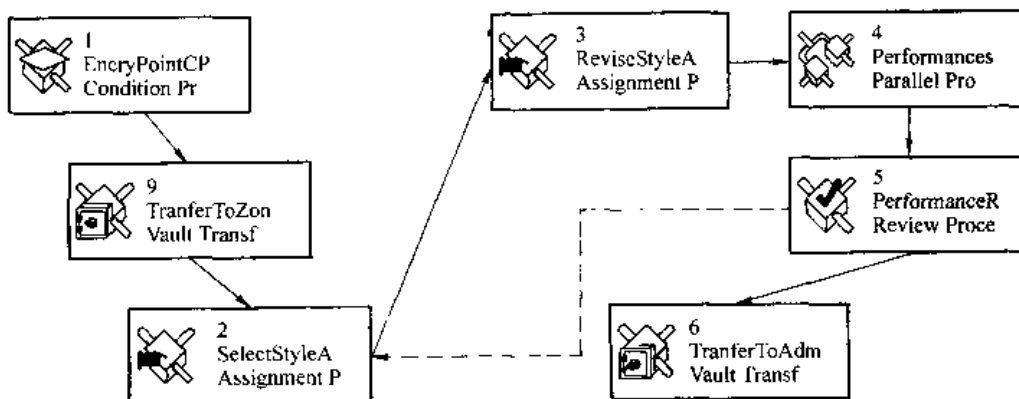


图 39.7-33 内饰件设计、评审及数据转换工作

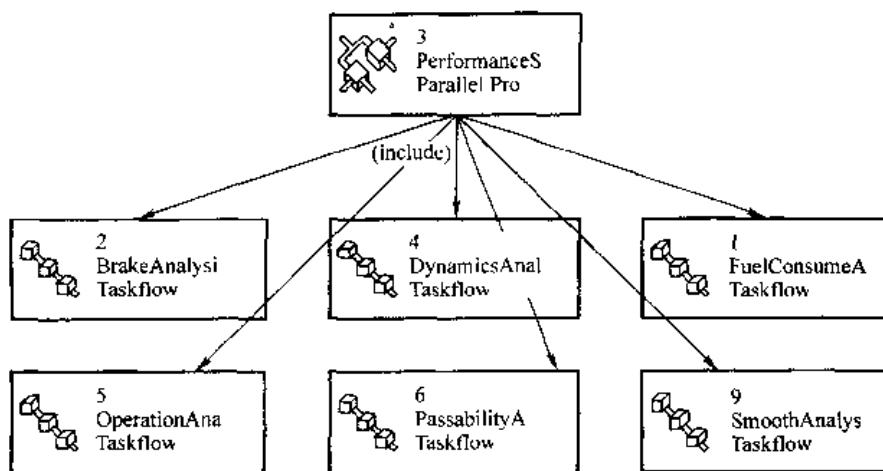


图 39.7-34 客车车身总体布置设计流程

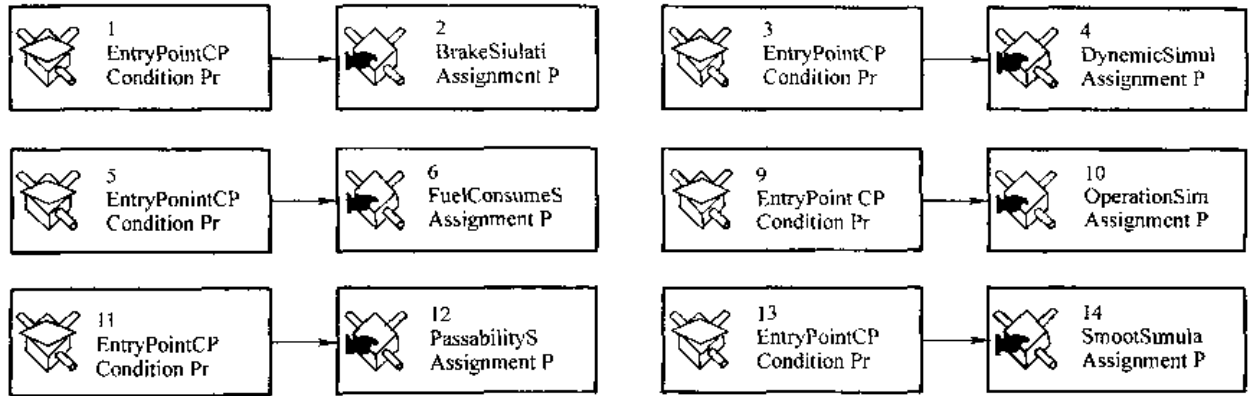


图 39.7-35 内饰设计任务分配

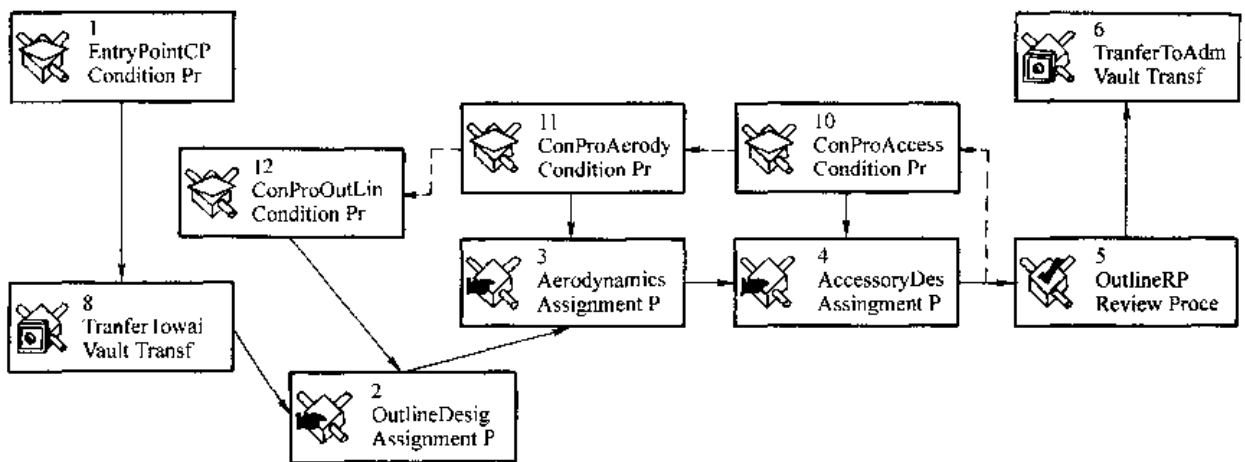


图 39.7-36 外形设计控制流程

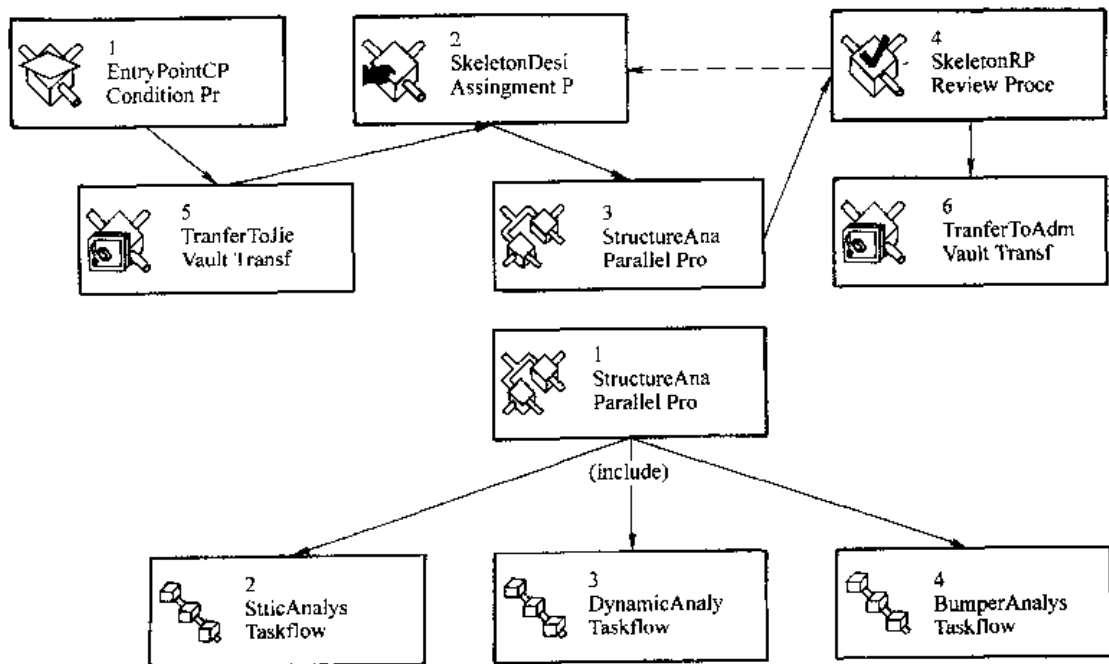


图 39.7-37 客车车身结构设计流程

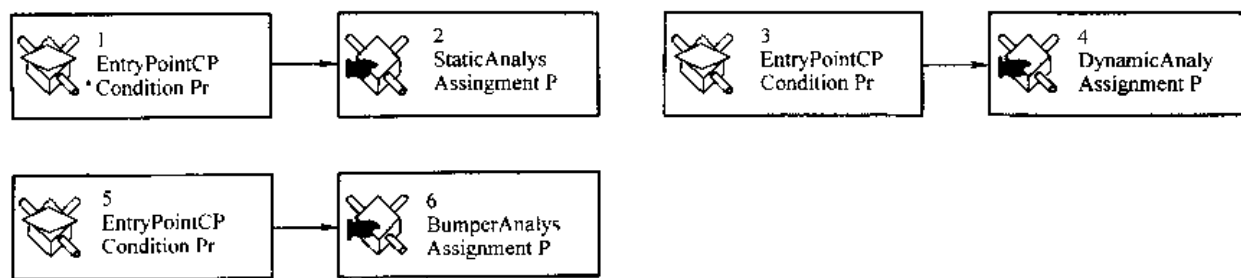


图 39.7-38 客车车身结构分析流程

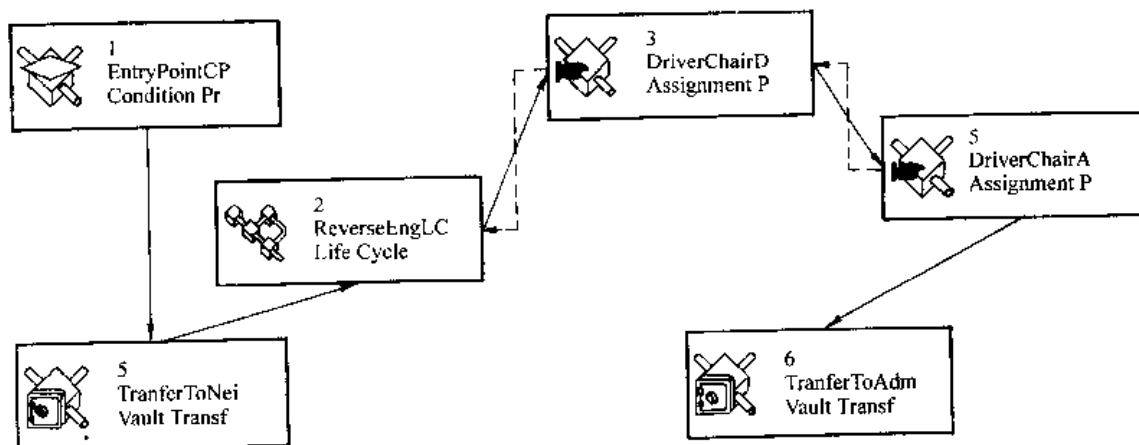


图 39.7-39 驾驶员座椅设计流程

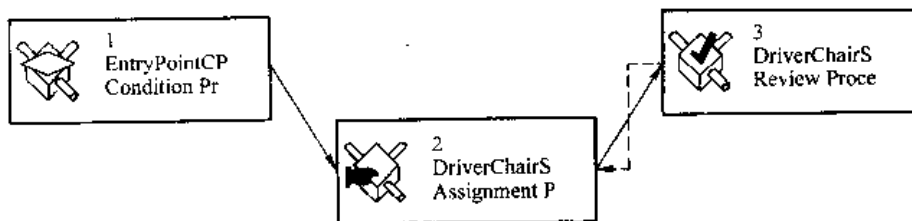


图 39.7-40 反求工程控制流程

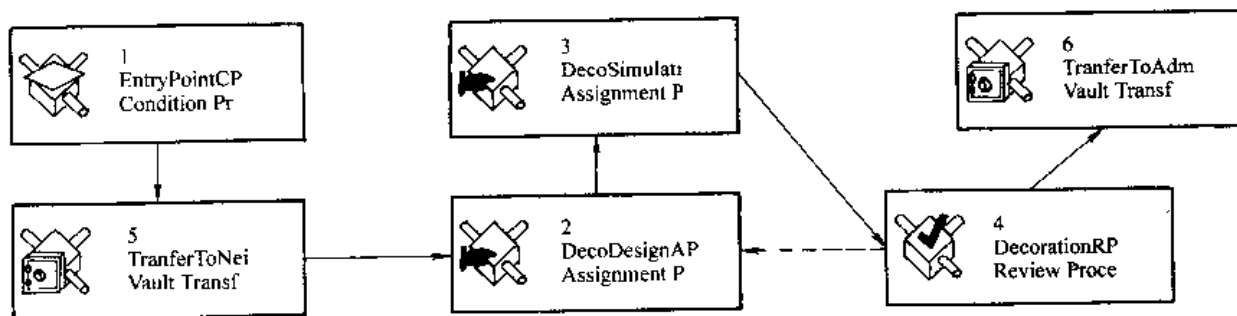


图 39.7-41 内饰设计流程

## 2.9 多反馈机制及实现

在 Metaphase 提供的流程管理功能中,一个 review 过程被拒绝后,只能将流程反馈到一个指定的过程。但在工厂的实际生产过程中,通常一个过程的反馈是不确定的,在不同的时候一个 review 过程被拒绝后可能要反馈到不同的过程,由此可见, Metaphase 提供的流程管理功能有着不完善的地方。

在本项目中,流程中涉及到的过程比较多,为了尽可能真实地反映工厂企业实际的生产过程,我们考虑引入多反馈机制。如果要实现多反馈,必须利用 Metaphase 提供的 Model 语言进行客户化开发,开发的过程介绍如下:

为了实现多反馈,我们决定在 Signoff 菜单中增加一个名为 Return to process 的菜单项,在签审操作中,当用户单击该菜单项时,会出现一个对话框,在这个对

对话框的列表框里,将显示出同一个流程中当前 review 过程前的所有 assignment 过程名称。这样,在这个列表框里,用户就可以看到他要反馈到的过程名称,单

击选取该过程,从而将流程反馈到该过程,增加的菜单如图 39-7-44 所示。

单击该菜单项后出现的对话框如图 39-7-45 所示。

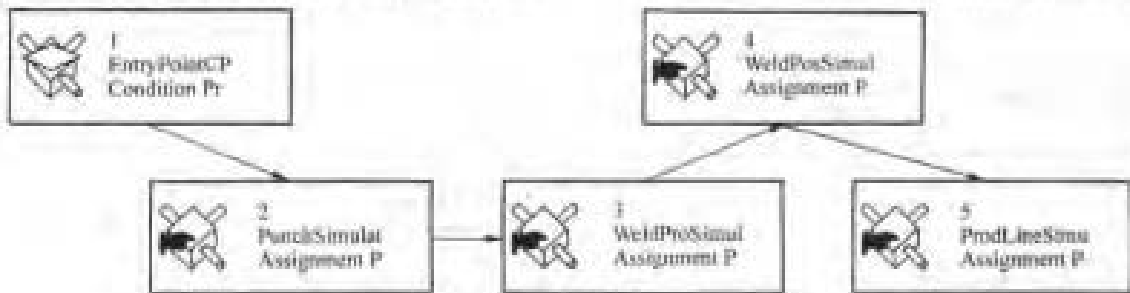


图 39-7-42 客车车身可制造性分析流程

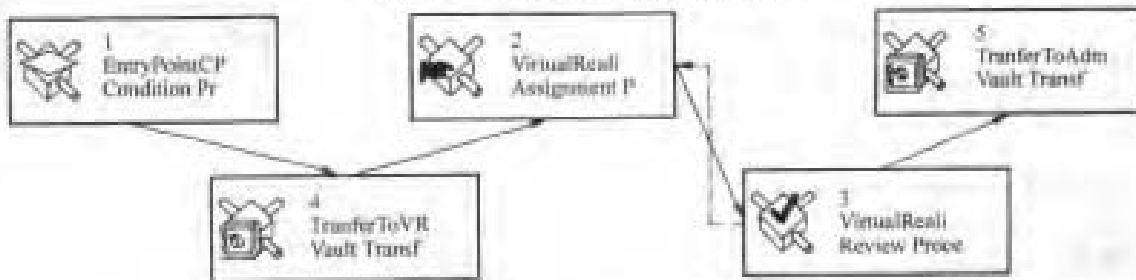


图 39-7-43 并行设计流程框流程图

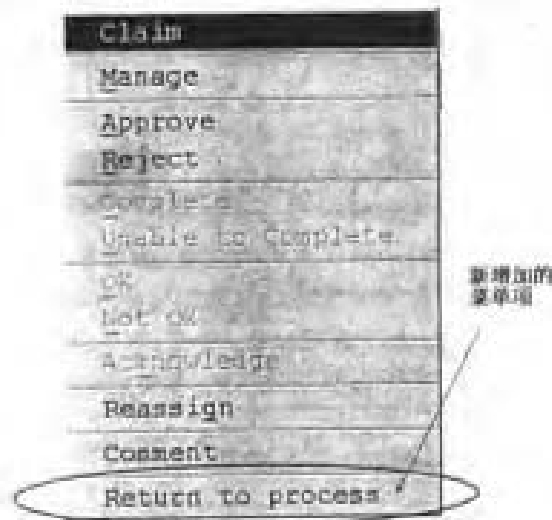


图 39-7-44 设计反馈活动菜单

在图中,右边对话框列表框里列出的是当前要拒绝的 review 过程前面已经流过的 Assignment 过程名称,当用户单击其中的某一个过程名称,再单击 OK 按钮后,该过程名称便会显示到左边对话框的文本框中,接着如果单击左边对话框中的 OK 按钮和 reject 菜单项,便可以将流程反馈到选取的 Assignment 过程上去。

利用 Metaphase 提供的 Model 语言可以实现上述菜单项的增加和对话框的出现。

首先找到 Signoff 菜单在 Metaphase 中对应的 Menu 名称为 LcmRevWrkSignoffMenu,接着自己定义了一个名为 ReturnProcessOptL 的 Option List 和名为 ReturnProcessOpt 的 Option,将 ReturnProcessOpt 附加到 ReturnProcessOptL,再将 ReturnProcessOptL 附加到 LcmRevWrkSignoffMenu 就可以出现新菜单了。

为了单击新增加的菜单后能出现对话框,首先定义了对话框类 Dreturn,再定义了该类的一个名为

ReturnProcess 的属性,接着定义一个名为 ReturnProcessSet 的数值集附加到该属性上。对话框就这样定义好了,并且在对话框上会出现一个复选框,单击该复选框能够出现一个列表框,该列表框里就显示了 ReturnProcessSet 数值集中的值,即当前 review 过程前面的 assignment 过程的名称。

为了能够使流程反馈到选取的 assignment 过程,

我们为每个有可能反馈到的 assignment 过程增加了一个条件过程(Condition Process),为该条件过程赋予的条件成立时,流程流到该条件过程对应的 assignment 过程,条件不成立时,流程流到下一个条件过程。条件示例如下所示:

Return To Design (0,c)<sub>1</sub> = \$ a. WhichProcess = 'yDesign Assignment';

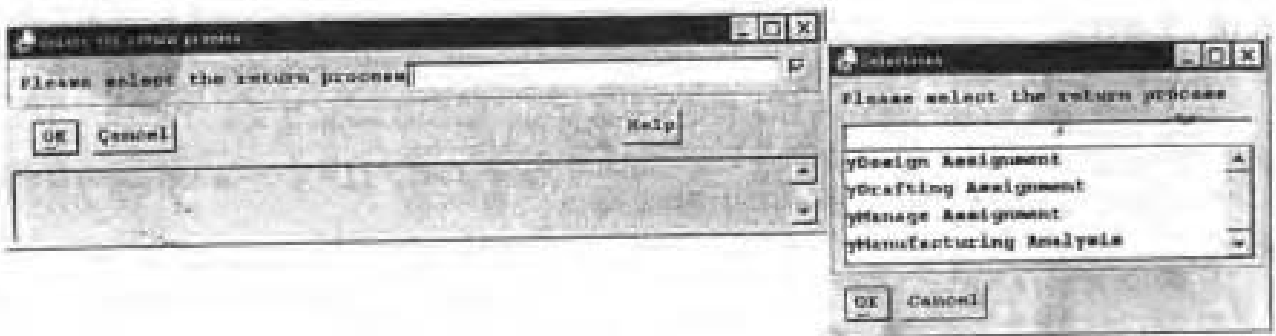


图 16.7-45 设计反馈对话框

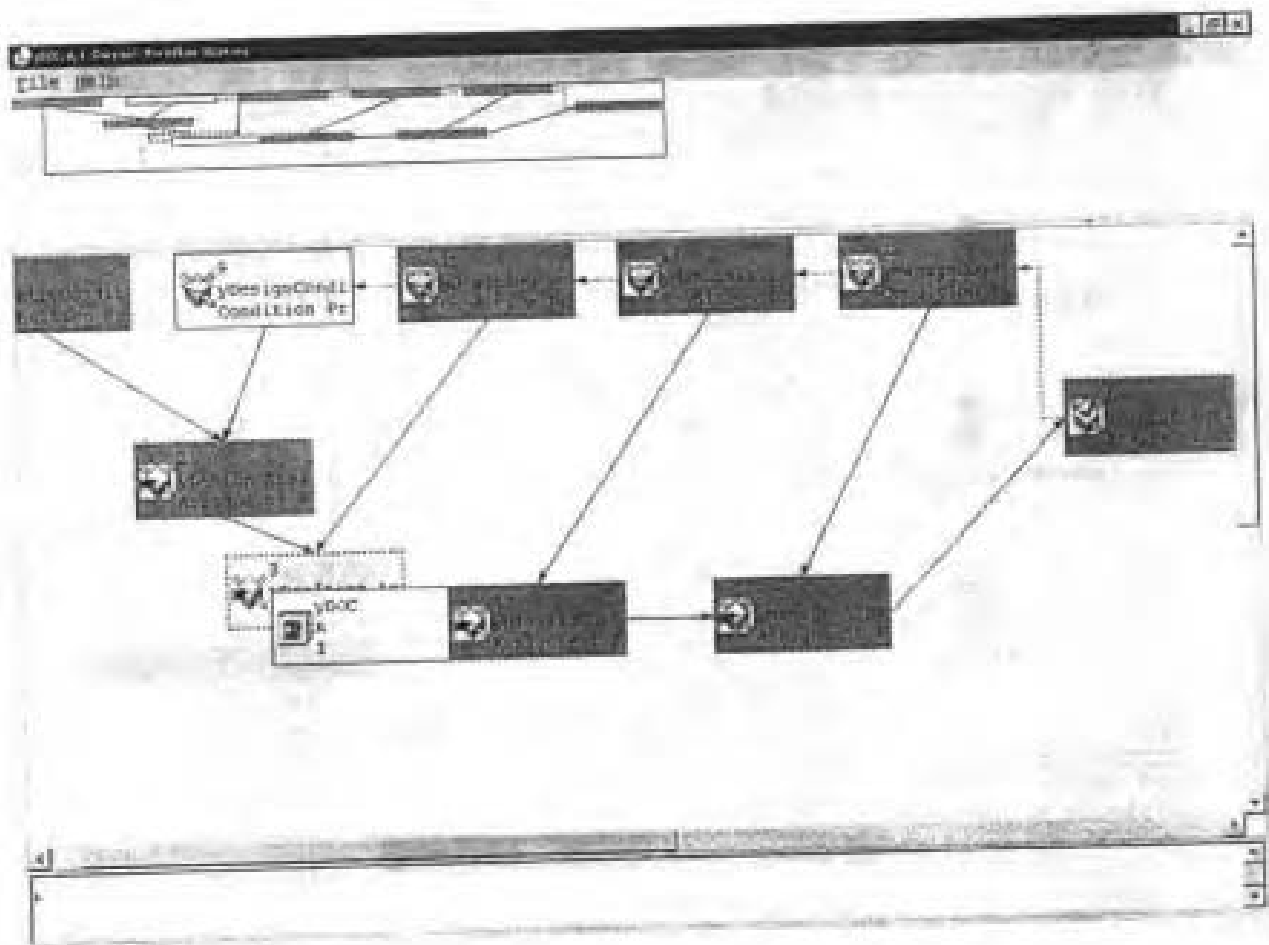


图 16.7-46 多反馈流程定义

多反馈的实现如图 39-7-46 所示。图中可以看到，当流程流到 yManage Review 过程要拒绝时，可以先选取要反馈到的 assignment 过程。在本例中，选取了 yDrafting Assignment 过程，则流程流经上面的 yManageCondition、yManufacturingCondition 条件过程到 yDraftingCondition 过程时，yDraftingCondition 条件过程对应的条件为真，流程便会流向 yDrafting Assignment 过程。

## 2.10 工具封装

### (1) 封装工具分类

- 1) 办公自动化软件 Microsoft Word7.0;  
Microsoft Visio4.1
- 2) 三维 CAD 造型软件 Pro/E2000I; Surface  
10.0
- 3) 程分析软件 Ansys5.5; Adams10.0
- 4) 并行设计漫游软件 RealX

### (2) 定义工具类

为应用工具定义一个新的工具类(该类以 Metaphase 中的 Tool 类作为父类)，并给该类的常量 DefaultCommandC 赋予一个可执行的应用程序名。

#### 定义应用文件类

1) 为与新工具类相关的每个文件类型定义一个数据类，该类的主要属性：

IconFamilyC 定义工具类的缺省图标对应的文件名；

FileNamePatternC 为方便注册定义缺省的文件命名方式；

LaunchToolC 定义当用户双击数据图标时被激活的工具类。

### (3) 定义用户界面

用于创建、查询特定的应用文件的界面分别如图 39-7-47、图 39-7-48 所示。

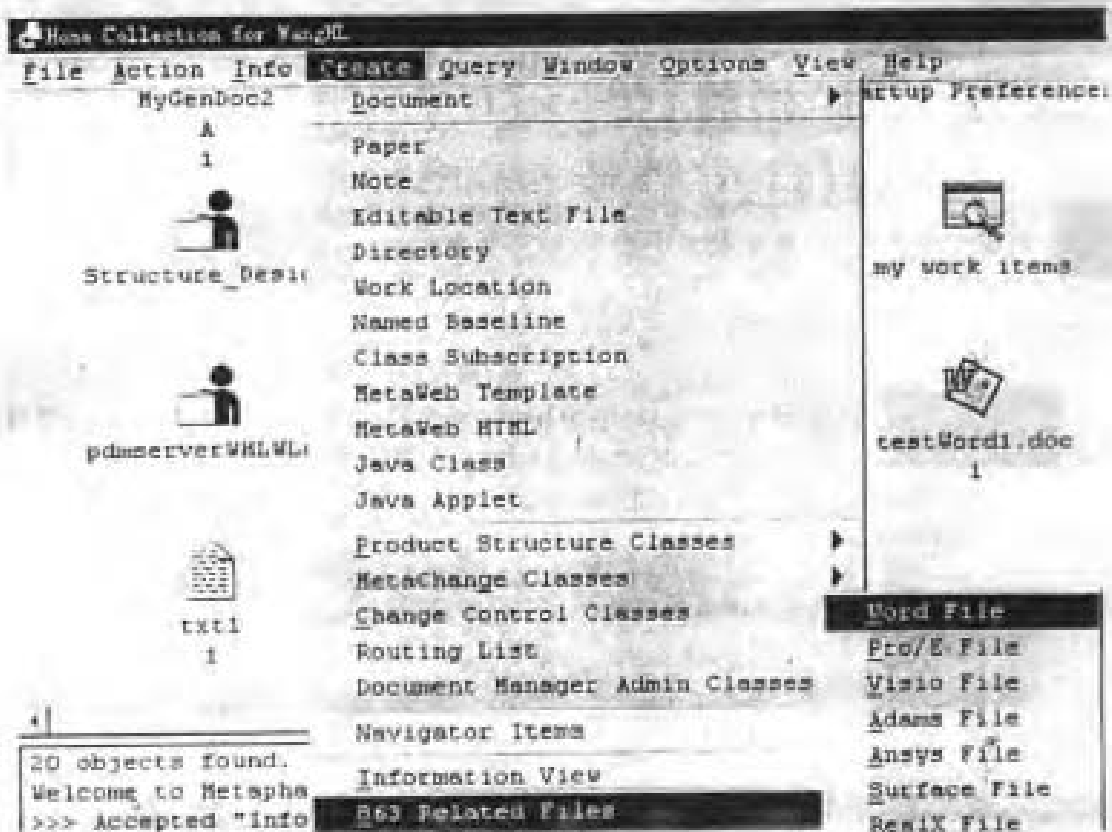


图 39-7-47 应用文件类创建界面





图 7-7-48 应用文件类查询界面

### 参考文献

- 1 Ed Miller. PDM today. Computer Aided Engineering, 1995, 14(2), 32~41
- 2 D. Burdick. Product Data Management, Enabling Enterprisewide Design Collaboration CIM Strategic Analysis Report, 8, Sept, 1995
- 3 王建民, 孙家广. 产品数据管理(PDM)技术及其应用. 电子展望与决策, 1997(5), 35~36
- 4 宋文博, 肖毅, 王德庆. 产品数据管理. 计算机应用研究, 1999(6), 11~12
- 5 朱志良, 庞士宗, 陆一平, 胡伟. 产品数据管理(PDM)技术的新思考. 机械设计与制造, 1999(1), 14~17
- 6 孟明辰, 韩向利. 并行设计. 北京, 机械工业出版社, 1999
- 7 孙莹, 陈松乔. 产品数据管理系统对企业数据信息的集成. 计算技术与自动化, 2000(3), 73~76
- 8 俞诗, 吕震, 方明伦, 吴国成, 郑玉培. 应用软件系统在 PDM 平台上的集成机制及其实现. 中国机械工程, 2000(12), 1379~1381
- 9 米小珍, 黄明, 刘晓冰. 并行开发中基于过程的产品数据管理. 组合机床与自动化加工技术, 2001(2), 44~47
- 10 冯开华, 李继明, 董秉枢. PDM 系统与群件系统的集成. 计算机辅助设计与图形学学报, 2001(4), 362~365
- 11 雷刚, 邹昌平. 产品数据管理技术概述及其发展趋势. 包装工程, 2001(4), 47~49
- 12 代颖, 马祖军. 基于 PDM 的企业 CAx/ERP 信息集成技术研究. 机电一体化, 2001(6), 16~19
- 13 李春平, 刘乃若, 郭鸣. 产品数据标准与 PDM. 北京, 清华大学出版社, 2002
- 14 任光胜, 许培硬. 产品数据管理技术的发展与应用. 组合机床与自动化加工技术, 2002(10), 67~71

# 第 40 篇 有限元法及其工程应用

主 编 林忠钦

编写人 王 华 韩 旭

有限元法又称有限单元法,是求解偏微分方程的一种有效的数值方法。有限元法发展至今天,已成为工程数值分析的有力工具。特别是在固体力学和结构分析的领域内,有限元法取得了巨大的进展,利用它已成

功地解决了一大批有重大意义的问题,很多通用程序和专用程序投入了实际应用,而且应用愈加广泛,已拓展到诸多领域。

## 第 1 章 弹性力学问题有限元法原理和表达式

### 1 广义坐标有限元法的一般格式

二维或三维连续体离散化为有限个单元的集合体,要求单元具有简单而规则的几何形状以便于计算。单元往往具有不同的形状和节点数,需要根据求解问题的类型、对计算精度的要求以及经济性等多方面的因素来选择合适的单元进行计算

#### 1.1 选择单元位移函数的一般原则

单元中的位移模式一般采用以广义坐标为待定参数的有限项多项式作为近似函数,有限项多项式选取的原则应考虑以下几点:

1) 广义坐标是由节点场变量确定的,因此它的个数应与节点自由度数相等。

2) 选取多项式时,常数项和坐标的一次项必须完备。位移模式中必须包含反映单元刚体位移和常应变的常数项和一次项。当划分的单元数趋于无穷时,单元缩小趋于一点,此时单元应变应趋于常应变。

3) 多项式的选取应由低阶到高阶,尽量选取完全多项式以提高单元的精度。若由于项数限制不能选取完全多项式时,选择的多项式应具有坐标的对称性。并且一个坐标方向的次数不应超过完全多项式的次数。

#### 1.2 广义坐标有限元的一般格式

1) 以广义坐标  $\{\beta\}$  为待定参数,给出单元内位移  $\{\delta\}$ ;

$$\{\delta\} = \{\phi\} \{\beta\} = N \{\delta\}^e$$

$$\{\phi\} = \begin{bmatrix} \varphi & 0 & \cdots \\ 0 & \varphi & \cdots \end{bmatrix}$$

其中  $\varphi = [1 \quad x \quad y \quad \cdots]$ ,  $N$  为形状函数矩阵,  $\{\delta\}^e$  为单元节点位移。

2) 用单元节点位移  $\{\delta\}^e$  表示单元应变和应力。

单元应变:

$$\{\epsilon\}^e = B \{\delta\}^e$$

单元应力:

$$\{\sigma\}^e = D \{\epsilon\}^e = DB \{\delta\}^e = S \{\delta\}^e$$

式中  $B$ ——几何矩阵;

$D$ ——弹性矩阵;

$S$ ——应力矩阵。

3) 节点载荷。

分布体积力的单元等效节点载荷

$$\{P\}_f^e = \iiint_{V_e} N^T \{f\} dx dy dz$$

分布面力的单元等效节点载荷

$$\{P\}_p^e = \int_S N^T \{q\} dS$$

初应力与初应变的单元等效节点载荷

$$\{P\}_{\epsilon_0}^e = \iiint_{V_e} B^T D \{\epsilon_0\} dx dy dz$$

$$\{P\}_{\sigma_0}^e = \iiint_{V_e} B^T \{\sigma_0\} dx dy dz$$

节点载荷列阵  $\{P\}$  可以写作

$$\begin{aligned} \{P\} &= \{P\}_f + \{P\}_p + \{P\}_{\epsilon_0} + \{P\}_{\sigma_0} + \{P\}_F \\ &= \sum_c G^T \{ \{P\}_f^e + \{P\}_p^e + \{P\}_{\epsilon_0}^e + \{P\}_{\sigma_0}^e \} + \{P\}_F \end{aligned}$$

式中  $\{P\}_F$ ——节点集中力列阵。

4) 用最小位能原理建立离散体系的节点平衡方程。

节点平衡方程

$$K \{\delta\} = \{P\}$$

其中整体刚度矩阵  $K = \sum_c G^T K^e G$ , 单元刚度矩阵  $K^e$

$$= \int_{V_e} B^T D B dV, G \text{ 为变换矩阵。}$$

### 2 平面问题三角形单元的有限元格式

典型的 3 节点三角形单元三个节点按逆时针编号, 设为  $i, j, m$ , 如图 40.1-1 所示。单元的分析公式见表 40.1-1。

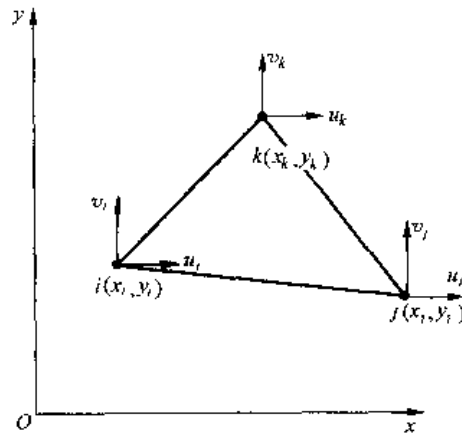


图 40.1-1 3 节点三角形单元

表 40.1-1 3 节点三角形单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [X_i \ Y_i \ X_j \ Y_j \ X_k \ Y_k]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u_i \ v_i \ u_j \ v_j \ u_k \ v_k]^T$
3	位移函数	$u$ $v$	$u = [N_i \ N_j \ N_k] \begin{Bmatrix} u_i \\ u_j \\ u_k \end{Bmatrix}, v = [N_i \ N_j \ N_k] \begin{Bmatrix} v_i \\ v_j \\ v_k \end{Bmatrix}$ $N_i = \frac{1}{2\Delta}(a_i + b_i x + c_i y), \text{ 其中 } \begin{cases} a_i = \begin{vmatrix} x_j & y_j \\ x_k & y_k \end{vmatrix} = x_j y_k - x_k y_j \\ b_i = -\begin{vmatrix} 1 & y_j \\ 1 & y_k \end{vmatrix} = y_k - y_j \\ c_i = \begin{vmatrix} 1 & x_j \\ 1 & x_k \end{vmatrix} = -x_j + x_k \end{cases} \quad (i, j, k)$ $\Delta = \frac{1}{2}(b_i c_j - b_j c_i)$
4	应变	$\epsilon^e$	$\epsilon^e = [\epsilon_x \ \epsilon_y \ \gamma_{xy}]^T = B\delta^e$ $B = \frac{1}{2\Delta} \begin{bmatrix} b_i & 0 & b_j & 0 & b_k & 0 \\ 0 & c_i & 0 & c_j & 0 & c_k \\ c_i & b_i & c_j & b_j & c_k & b_k \end{bmatrix}$
5	应力	$\sigma^e$	$\sigma^e = [\sigma_x \ \sigma_y \ \tau_{xy}]^T = D\epsilon^e = DB\delta^e = S\delta^e$ <p>平面应力问题</p> $D = \frac{E}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1-\mu}{2} \end{bmatrix} \quad S = DB$ <p>平面应变问题 将 <math>D</math> 和 <math>S</math> 式中的 <math>E</math> 和 <math>\mu</math> 分别换成 <math>E/(1-\mu^2)</math> 和 <math>\mu/(1-\mu)</math></p>
6	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = B^T D B t \Delta = \begin{bmatrix} [k_{ii}^e] & [k_{ij}^e] & [k_{ik}^e] \\ [k_{ji}^e] & [k_{jj}^e] & [k_{jk}^e] \\ [k_{ki}^e] & [k_{kj}^e] & [k_{kk}^e] \end{bmatrix}$ <p>平面应力问题</p> $k_{rs}^e = \frac{Et}{4(1-\mu^2)\Delta} \begin{bmatrix} b_r b_s - \frac{1-\mu}{2} c_r c_s & \mu b_r c_s + \frac{1-\mu}{2} b_r c_s \\ \mu c_r b_s + \frac{1-\mu}{2} b_r c_s & c_r c_s + \frac{1-\mu}{2} b_r b_s \end{bmatrix} \quad (r=i, j, k; s=i, j, k)$ <p>平面应变问题 将 <math>D</math> 和 <math>S</math> 式中的 <math>E</math> 和 <math>\mu</math> 分别换成 <math>E/(1-\mu^2)</math> 和 <math>\mu/(1-\mu)</math></p>

(续)

序号	类别	符号	公式
7	总体刚度矩阵	$K$	$K = \sum_{e=1}^m K^e$ $k^e = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \dots & \dots & k_{ii}^e & \dots & k_{ij}^e & \dots & k_{ik}^e & \dots & \dots \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \dots & \dots & k_{ji}^e & \dots & k_{jj}^e & \dots & k_{jk}^e & \dots & \dots \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \dots & \dots & k_{ki}^e & \dots & k_{kj}^e & \dots & k_{kk}^e & \dots & \dots \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$
8	整体平衡方程		$K\delta = P$ $\delta = [u_1 \ v_1 \ u_2 \ v_2 \ \dots \ u_{2n} \ v_{2n}]^T$ $P = [X_1 \ Y_1 \ X_2 \ Y_2 \ \dots \ X_{2n} \ Y_{2n}]^T$
9	节点等效载荷		作用在 $ij$ 边上一点与 $x$ 轴平行的集中力 $P$ 作用点距节点 $i$ 为 $l_i$ , 距节点 $j$ 为 $l_j$ $P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{P}{l} [l_j \ 0 \ l_i \ 0 \ 0 \ 0]^T$
			作用在 $ij$ 边的均布侧压 $q$ $P_q^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{1}{2} qt [y_i - y_j \ x_j - x_i \ y_i - y_j \ x_j - x_i \ 0 \ 0]^T$
			作用在 $ij$ 边的与 $x$ 轴平行的均布面力 $q$ $P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{1}{2} qt [1 \ 0 \ 1 \ 0 \ 0 \ 0]^T$
			作用在 $ij$ 边的与 $x$ 轴平行的三角形分布面力 $q$ $\begin{cases} q_i = q \\ q_j = 0 \end{cases} P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{1}{2} qt \left[ \frac{2}{3} \ 0 \ \frac{1}{3} \ 0 \ 0 \ 0 \right]^T$
			体积力 $f$ $f = \begin{Bmatrix} f_x \\ f_y \end{Bmatrix} P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{1}{3} t \Delta [f_x \ f_y \ f_x \ f_y \ f_x \ f_y]^T$
备注:	$N_i, N_j, N_k$ —单元的形状函数, 简称形函数; $\Delta$ —三角形单元的面积; $B$ —几何矩阵, 其各元素均为常量; $D$ —弹性矩阵; $S$ —应力矩阵; $t$ —单元厚度; $m$ —单元数; $K^e$ — $K^e$ 扩充成的 $2n \times 2n$ 阶单元贡献矩阵; $n$ —节点数; $\delta$ —所有节点的位移矢量; $P$ —所有节点的节点力矢量, 为直接作用于节点上的集中载荷矢量与等效节点载荷矢量之和。		

### 3 平面问题矩形单元的有限元格式

边长为  $2a$  及  $2b$  的矩形单元, 取 4 个角点作为节点, 用局部节点编码  $i \sim l$  表示, 如图 40.1-2 所示。以单元形心  $O$  为原点建立局部坐标  $xOy$ , 并以平行于两边的两个对称轴作为  $x$  轴和  $y$  轴。为运算方便, 引入无量纲局部坐标  $\xi, \eta$ , 且

$$\xi = \frac{x}{a}, \eta = \frac{y}{b}$$

单元的分析公式见表 41.1-2。

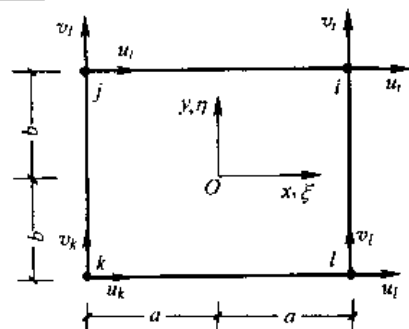


图 40.1-2 4 节点矩形单元

表 40.1-2 4 节点矩形单元的分析公式

序号	类别	符号	公式
1	节点力 矢量	$F^e$	$F^e = [X_i, Y_i, X_j, Y_j, X_k, Y_k, X_l, Y_l]^T$
2	节点位 移矢量	$\delta^e$	$\delta^e = [u_i, v_i, u_j, v_j, u_k, v_k, u_l, v_l]^T$
3	位移 函数	$u, v$	$u = [N_i, N_j, N_k, N_l]^T \begin{Bmatrix} u_i \\ u_j \\ u_k \\ u_l \end{Bmatrix}, v = [N_i, N_j, N_k, N_l]^T \begin{Bmatrix} v_i \\ v_j \\ v_k \\ v_l \end{Bmatrix}$ $N_i = \frac{1}{4}(1-\xi)(1-\eta), N_j = \frac{1}{4}(1+\xi)(1-\eta)$ $N_k = \frac{1}{4}(1+\xi)(1-\eta), N_l = \frac{1}{4}(1-\xi)(1-\eta)$
4	应变	$\epsilon^e$	$\epsilon^e = [\epsilon_x, \epsilon_y, \gamma_{xy}]^T = B\delta^e$ $B = \frac{1}{4ub} \begin{bmatrix} -(b-y) & 0 & b-y & 0 & b+y & 0 & -(b-y) & 0 \\ 0 & (a-x) & 0 & -(a-x) & 0 & a-x & 0 & a-x \\ -(a-x) & -(b-y) & -(a+x) & b-y & a+x & b+y & a-x & -(b+y) \end{bmatrix}$
5	应力	$\sigma$	$\sigma = [\sigma_x, \sigma_y, \tau_{xy}]^T = D\epsilon^e = DB\delta^e = S\delta^e$ <p>平面应力问题</p> $D = \frac{E}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1-\mu}{2} \end{bmatrix} \quad S = DB$ <p>平面应变问题 将 <math>D</math> 和 <math>S</math> 式中的 <math>E</math> 和 <math>\mu</math> 分别换成 <math>E/(1-\mu^2)</math> 和 <math>\mu/(1-\mu)</math></p>
6	单元刚 度矩阵	$K^e$	$K^e = \begin{bmatrix} k_{11}^e & k_{12}^e & k_{13}^e & k_{14}^e \\ k_{12}^e & k_{22}^e & k_{23}^e & k_{24}^e \\ k_{13}^e & k_{23}^e & k_{33}^e & k_{34}^e \\ k_{14}^e & k_{24}^e & k_{34}^e & k_{44}^e \end{bmatrix}$ <p>平面应力问题</p> $k_{rs}^e = \frac{E}{4ab(1-\mu^2)} \begin{bmatrix} \frac{b}{a} \xi_r \xi_s \left( 1 + \frac{1}{3} \eta_r \eta_s \right) + \frac{1-\mu}{2} \frac{b}{a} \eta_r \eta_s \left( 1 + \frac{1}{3} \xi_r \xi_s \right) & \mu \xi_r \eta_s + \frac{1-\mu}{2} \eta_r \xi_s \\ \mu \xi_r \eta_s + \frac{1-\mu}{2} \eta_r \xi_s & \frac{a}{b} \eta_r \eta_s \left( 1 - \frac{1}{3} \xi_r \xi_s \right) + \frac{1-\mu}{2} \frac{a}{b} \xi_r \xi_s \left( 1 + \frac{1}{3} \eta_r \eta_s \right) \end{bmatrix}$ <p>(<math>r=s=1, 2, 3, 4</math>; <math>s=1, 2, 3, 4</math>)</p> <p>平面应变问题 将 <math>D</math> 和 <math>S</math> 式中的 <math>E</math> 和 <math>\mu</math> 分别换成 <math>E/(1-\mu^2)</math> 和 <math>\mu/(1-\mu)</math></p>
7	节点等 效载荷		<p>作用在 <math>ij</math> 边上一点与 <math>x</math> 轴平行的集中力 <math>P</math> 作用点距节点 <math>i</math> 为 <math>l</math>, 距节点 <math>j</math> 为 <math>l_j</math></p> $P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \\ P_l \end{Bmatrix} = \frac{P}{a} [l, 0, l_j, 0, 0, 0, 0, 0]^T$ <p>作用在 <math>ij</math> 边的与 <math>x</math> 轴平行的均布面力 <math>q</math></p> $P^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \\ P_l \end{Bmatrix} = qat [1, 0, 1, 0, 0, 0, 0, 0]^T$

(续)

序号	类别	符号	公式
7	节点等效载荷		作用在 $ij$ 边的与 $x$ 轴平行的三角形分布面力 $q$ $\begin{Bmatrix} q_r = q \\ q_\theta = 0 \end{Bmatrix} \quad P^e = \begin{Bmatrix} P_1 \\ P_2 \\ P_3 \end{Bmatrix} = qat \begin{bmatrix} 2 & 0 & 1 \\ 3 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}^T$
			体积力 $f$ $f = \begin{Bmatrix} f_r \\ f_\theta \\ f_z \end{Bmatrix} \quad P^v = \begin{Bmatrix} P_1 \\ P_2 \\ P_3 \end{Bmatrix} = tab \begin{bmatrix} f_r & f_\theta & f_r & f_\theta & f_z & f_\theta & f_r & f_\theta \end{bmatrix}^T$
备注			$N_i, N_j, N_k, N_l$ —单元的形状函数, 简称形函数; $B$ —几何矩阵; $D$ —弹性矩阵; $S$ —应力矩阵; $t$ —单元厚度; $K^e$ — $K^*$ 扩充成的 $2n \times 2n$ 阶单元贡献矩阵; $n$ —节点数; $\delta$ —所有节点的位移矢量; $P$ —所有节点的节点力矢量, 为直接作用于节点上的集中载荷矢量与等效节点载荷矢量之和。

#### 4 轴对称问题的有限元格式

在轴对称问题中, 通常采用圆周坐标  $(r, \theta, z)$ , 以  $z$  轴作为对称轴, 所有应力、应变和位移都与  $\theta$  方向无关, 只是  $r$  和  $z$  的函数, 且  $\theta$  方向的位移  $v$  等于零。将旋转体离散成有限个轴对称的圆环单元, 这些圆环

单元与  $rz$  平面正交的截面可以有不同的形状(三角形或矩形等), 各单元在  $rz$  平面内形成网格, 可象平面问题一样处理。如图 40.1-3 所示, 图中  $z$  轴为对称轴, 轴对称问题三角形单元上每个节点有两个节点力  $R, Z$  和两个节点位移  $u, w$ , 单元的分析公式见表 40.1-3。

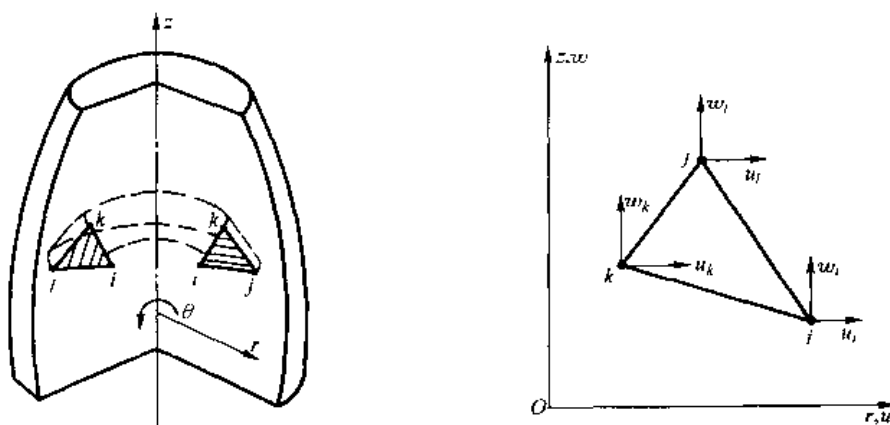


图 40.1-3 3 节点三角形环状单元及其  $rz$  截面

表 40.1-3 轴对称问题三角形单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [R_i \ Z_i \ R_j \ Z_j \ R_k \ Z_k]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u_i \ w_i \ u_j \ w_j \ u_k \ w_k]^T$
3	位移函数	$u$ $w$	$u = [N_i \ N_j \ N_k] \begin{Bmatrix} u_i \\ u_j \\ u_k \end{Bmatrix}, w = [N_i \ N_j \ N_k] \begin{Bmatrix} w_i \\ w_j \\ w_k \end{Bmatrix}$ $N_i = \frac{1}{2\Delta} (a_i + b_i r + c_i z), \text{ 其中 } \begin{cases} a_i = \begin{vmatrix} r_j & z_j \\ r_k & z_k \end{vmatrix} = r_j z_k - r_k z_j \\ b_i = -\begin{vmatrix} 1 & x_j \\ 1 & x_k \end{vmatrix} = z_j - z_k \\ c_i = \begin{vmatrix} 1 & r_j \\ 1 & r_k \end{vmatrix} = -r_j + r_k \end{cases} \quad (i, j, k)$ $\Delta = \frac{1}{2} (b_i c_i - b_j c_j)$

(续)

序号	类别	符号	公式
4	应变	$\epsilon^e$	$\epsilon^e = [\epsilon_r \quad \epsilon_\theta \quad \epsilon_z \quad \gamma_{rz}]^T = \mathbf{B}\delta^e$ $\mathbf{B} = \frac{1}{2\Delta} \begin{bmatrix} b_i & 0 & b_j & 0 & b_k & 0 \\ g_i & 0 & g_j & 0 & g_k & 0 \\ 0 & c_i & 0 & c_j & 0 & c_k \\ c_i & b_i & c_j & b_j & c_k & b_k \end{bmatrix}$ $g_s = \frac{a_s}{r} + b_s + \frac{c_s z}{r} \quad (s=i, j, k)$
5	应力	$\sigma$	$\sigma = [\sigma_r \quad \sigma_\theta \quad \sigma_z \quad \tau_{rz}]^T = \mathbf{D}\epsilon^e = \mathbf{D}\mathbf{B}\delta^e = \mathbf{S}\delta^e$ $\mathbf{D} = \frac{E(1-\mu)}{(1+\mu)(1-2\mu)} \begin{bmatrix} 1 & \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 0 \\ \frac{\mu}{1-\mu} & 1 & \frac{\mu}{1-\mu} & 0 \\ \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 1 & 0 \\ 0 & 0 & 0 & \frac{1-2\mu}{2(1-\mu)} \end{bmatrix} \quad \mathbf{S} = \mathbf{D}\mathbf{B}$
6	单元刚度矩阵	$\bar{\mathbf{K}}^e$	$\bar{\mathbf{K}}^e = \begin{bmatrix} k_{ii}^e & k_{ij}^e & k_{ik}^e \\ k_{ji}^e & k_{jj}^e & k_{jk}^e \\ k_{ki}^e & k_{kj}^e & k_{kk}^e \end{bmatrix}$ $k_{mn}^e = \frac{\pi E(1-\mu)r}{2(1+\mu)(1-2\mu)} \times \begin{bmatrix} b_m b_n + \bar{g}_m \bar{g}_n + A_1(b_m \bar{g}_n + b_n \bar{g}_m) + A_2 c_m c_n & A_1(b_m c_n + \bar{g}_m c_n) + A_2 c_m b_n \\ A_1(c_m b_n - c_m \bar{g}_n) + A_2 b_m c_n & c_m c_n - A_2 b_m b_n \end{bmatrix}$ $(m=i, j, k; s=i, j, k)$ $\bar{g}_n = \frac{c_n}{r} + b_n + \frac{c_n z}{r} \quad (n=i, j, k)$ $\bar{r} = \frac{1}{3}(r_i + r_j + r_k), \bar{z} = \frac{1}{3}(z_i + z_j + z_k)$ $A_1 = \frac{\mu}{1-\mu}, A_2 = \frac{1-2\mu}{2(1-\mu)}$
7	节点等效载荷		<p>作用在 <math>ij</math> 边上一点与 <math>r</math> 轴平行的轴对称分布力 <math>q</math> 作用点距节点 <math>i</math> 为 <math>l</math>, 距节点 <math>j</math> 为 <math>l</math>,</p> $\mathbf{P}^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{2\pi r q}{l} [l, 0, l, 0, 0, 0]^T$ <p>体积力 <math>f</math></p> $\mathbf{f} = \begin{Bmatrix} f_r \\ f_z \end{Bmatrix} \quad \mathbf{P}^e = \begin{Bmatrix} P_i \\ P_j \\ P_k \end{Bmatrix} = \frac{2\pi r \Delta}{3} [f_r \quad f_z \quad f_r \quad f_z \quad f_r \quad f_z]^T$
备注	$N_i, N_j, N_k$ —单元的形函数, 简称形函数; $\Delta$ —三角形单元的面积; $\mathbf{B}$ —几何矩阵; $\mathbf{D}$ —弹性矩阵; $\mathbf{S}$ —应力矩阵。		

### 5 空间问题的有限元格式

弹性力学空间问题与平面问题类似, 先将空间结构离散成有限个三维单元 (通常为四面体单元或六面体单元), 各单元通过有限个节点相连, 求解步骤与平面问题相同。

空间四面体单元, 节点  $i, j, k, l$  按右手法则编号, 即从节点  $l$  看, 节点  $i, j, k$  为逆时针方向如图 40.1-4 所示。每个节点有三个节点力和节点位移, 单元的分析公式见表 40.1-4。

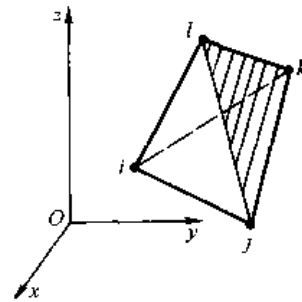


图 40.1-4 四面体单元



表 40.1-4 四面体单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [X_i, Y_i, Z_i, X_j, Y_j, Z_j, X_k, Y_k, Z_k, X_l, Y_l, Z_l]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u_i, v_i, w_i, u_j, v_j, w_j, u_k, v_k, w_k, u_l, v_l, w_l]^T$
3	位移函数	$u, v, w$	$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = \begin{bmatrix} N_i & 0 & 0 & -N_j & 0 & 0 & N_k & 0 & 0 & -N_l & 0 & 0 \\ 0 & N_i & 0 & 0 & -N_j & 0 & 0 & N_k & 0 & 0 & -N_l & 0 \\ 0 & 0 & N_i & 0 & 0 & -N_j & 0 & 0 & N_k & 0 & 0 & -N_l \end{bmatrix} \delta^e$ $N_r = \frac{1}{6V}(a_r - b_r x - c_r y + d_r z) \quad (r=i, j, k, l)$ <p>四面体体积</p> $V = \frac{1}{6} \begin{vmatrix} 1 & x_i & y_i & z_i \\ 1 & x_j & y_j & z_j \\ 1 & x_k & y_k & z_k \\ 1 & x_l & y_l & & z_l \end{vmatrix}$ $a_i = \begin{vmatrix} x_j & y_j & z_j \\ x_k & y_k & z_k \\ x_l & y_l & z_l \end{vmatrix}, b_i = \begin{vmatrix} 1 & y_j & z_j \\ 1 & y_k & z_k \\ 1 & y_l & z_l \end{vmatrix}, c_i = \begin{vmatrix} x_j & 1 & z_j \\ x_k & 1 & z_k \\ x_l & 1 & z_l \end{vmatrix}, d_i = \begin{vmatrix} x_j & y_j & 1 \\ x_k & y_k & 1 \\ x_l & y_l & 1 \end{vmatrix} \quad (i, j, k, l)$
4	应变	$\epsilon^e$	$\epsilon^e = [\epsilon_x, \epsilon_y, \epsilon_z, \gamma_{xy}, \gamma_{yz}, \gamma_{zx}]^T = B^T \delta^e$ $B = [B_i, -B_j, B_k, -B_l]$ $B_r = \frac{1}{6V} \begin{bmatrix} h_r & 0 & 0 \\ 0 & c_r & 0 \\ 0 & 0 & d_r \\ c_r & b_r & 0 \\ 0 & d_r & c_r \\ d_r & 0 & b_r \end{bmatrix} \quad (r=i, j, k, l)$
5	应力	$\sigma^e$	$\sigma^e = [\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{yz}, \tau_{zx}]^T = D \epsilon^e = DB^T \delta^e = S \delta^e$ $D = \frac{E(1-\mu)}{(1-\mu)(1-2\mu)} \begin{bmatrix} 1 & \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 0 & 0 & 0 \\ \frac{\mu}{1-\mu} & 1 & \frac{\mu}{1-\mu} & 0 & 0 & 0 \\ \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\mu}{2(1-\mu)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\mu}{2(1-\mu)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\mu}{2(1-\mu)} \end{bmatrix} \quad S = DB$
6	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \begin{bmatrix} k_{ii}^e & -k_{ij}^e & k_{ik}^e & -k_{il}^e \\ -k_{ji}^e & k_{jj}^e & -k_{jk}^e & k_{jl}^e \\ k_{ki}^e & -k_{kj}^e & k_{kk}^e & -k_{kl}^e \\ -k_{li}^e & k_{lj}^e & -k_{lk}^e & k_{ll}^e \end{bmatrix}$ $k_{rs}^e = \frac{E(1-\mu)}{36(1+\mu)(1-2\mu)} \times \begin{bmatrix} b_r b_s + A_2(c_r c_s + d_r d_s) & A_1 b_r c_s + A_2 c_r b_s & A_1 b_r d_s + A_2 d_r b_s \\ A_1 c_r b_s + A_2 b_r c_s & c_r c_s + A_2(b_r b_s + d_r d_s) & A_1 c_r d_s + A_2 d_r c_s \\ A_1 d_r b_s + A_2 b_r d_s & A_1 d_r c_s + A_2 c_r d_s & d_r d_s + A_2(b_r b_s + c_r c_s) \end{bmatrix}$ $(r=i, j, k, l; s=i, j, k, l)$ $A_1 = \frac{\mu}{1-\mu}, A_2 = \frac{1-2\mu}{2(1-\mu)}$
7	节点等效载荷		<p>作用在边界表面 <math>ijk</math> 的线性分布的面力</p> $p = [p_x, p_y, p_z]^T$ <p>在节点 <math>i, j, k</math> 上的集度分别为 <math>[p_x^i, p_y^i, p_z^i]^T, [p_x^j, p_y^j, p_z^j]^T, [p_x^k, p_y^k, p_z^k]^T</math>, 节点载荷如下:</p> $P_{ip} = \frac{1}{6} A_{ijk} \left( p_x^i + \frac{1}{2} p_x^j + \frac{1}{2} p_x^k \right) \quad (i, j, k)$ <p>式中: <math>A_{ijk}</math> 为边界表面 <math>ijk</math> 的面积。</p> <p>体积力 <math>f</math></p> $f = \begin{Bmatrix} f_x \\ f_y \\ f_z \end{Bmatrix} \quad P_{il} = \begin{Bmatrix} P_{ix} \\ P_{iy} \\ P_{iz} \end{Bmatrix} = \frac{V}{4} \begin{Bmatrix} f_x \\ f_y \\ f_z \end{Bmatrix} \quad (i, j, k, l)$
备注	$N_i, N_j, N_k, N_l$ —单元的形函数, 简称形函数; $B$ —几何矩阵; $D$ —弹性矩阵; $S$ —应力矩阵。		

## 第 2 章 等参单元与数值积分

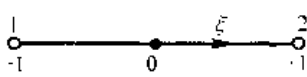
### 1 一维等参单元

利用自然坐标(在一维问题中可称为长度坐标)构造一维等参单元的形状函数。采用坐标  $\xi$ , 且  $-1 \leq \xi \leq +1$ , 如图 40.2-1 所示。自然坐标系中的直线单元称为母单元。

1) 线性单元(2 节点)

$$N_1 = \frac{1-\xi}{2}, N_2 = \frac{1+\xi}{2}$$

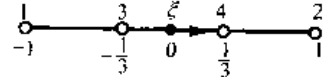
2) 二次单元(3 节点)



a)



b)



c)

图 40.2-1 一维母单元

a) 线性单元 b) 二次单元 c) 三次单元

通过坐标变换在整体坐标系中得到不同长度和形状的曲线单元, 称为子单元。在整体坐标系中, 一维子单元内任意一点的坐标用形函数表示如下:

$$x = \sum N_i x_i = N_1 x_1 + N_2 x_2 + \dots$$

$$y = \sum N_i y_i = N_1 y_1 + N_2 y_2 + \dots$$

其中  $N_i$  是用局部坐标表示的形函数;  $(x, y)$  是节点  $i$  的整体坐标。

### 2 二维等参单元

1) 三角形单元 三角形单元的形函数用面积坐标表示, 面积坐标与直角坐标的转换关系如图 40.2-2 所示。常用三角形平面单元的形函数见表 40.2-1。

$$N_1 = -\frac{(1-\xi)\xi}{2}, N_2 = \frac{(1+\xi)\xi}{2}, N_3 = 1-\xi^2$$

3) 三次单元(4 节点)

$$N_1 = \frac{(1-\xi)(9\xi^2-1)}{16}, N_2 = \frac{(1+\xi)(9\xi^2-1)}{16}$$

$$N_3 = \frac{9(1-\xi^2)(1-3\xi)}{16}, N_4 = \frac{9(1-\xi^2)(1+3\xi)}{16}$$

利用拉格朗日多项式, 一维形函数可统一写成如下形式:

下形式:

$$N_i = \frac{(\xi-\xi_1)(\xi-\xi_2)\dots(\xi-\xi_{i-1})(\xi-\xi_{i+1})\dots(\xi-\xi_n)}{(\xi_i-\xi_1)(\xi_i-\xi_2)\dots(\xi_i-\xi_{i-1})(\xi_i-\xi_{i+1})\dots(\xi_i-\xi_n)}$$

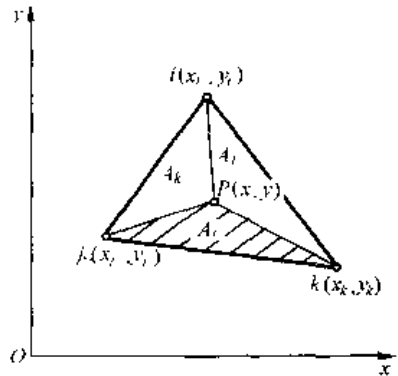


图 40.2-2 面积坐标与直角坐标的转换关系

表 40.2-1 常用三角形平面单元的形函数

单元名称	单元(母单元)形态	自由度	形函数
3 节点三角形平面元		$u, v$	$N_1 = L_1$ $N_2 = L_2$ ( $L_i$ 为面积坐标) $N_3 = L_3$
6 节点三角形平面元		$u, v$	$N_1 = (2L_1 - 1)L_1$ $N_2 = 4L_2L_3$ $N_3 = (2L_3 - 1)L_3$ $N_4 = 4L_1L_2$ $N_5 = (2L_2 - 1)L_2$ $N_6 = 4L_3L_1$

(续)

单元名称	单元(母单元)形态	自由度	形函数
10 节点三角形平面元		$u, v$	$N_1 = \frac{1}{2}(3L_1 - 1)(3L_1 - 2)L_1, N_2 = \frac{9}{2}L_1L_2(3L_1 - 1)$ $N_3 = \frac{9}{2}L_1L_2(3L_2 - 1), N_4 = \frac{1}{2}(3L_2 - 1)(3L_2 - 2)L_2$ $N_5 = \frac{9}{2}L_2L_3(3L_2 - 1), N_6 = \frac{9}{2}L_3L_2(3L_3 - 1)$ $N_7 = \frac{1}{2}(3L_3 - 1)(3L_3 - 2)L_3, N_8 = \frac{9}{2}L_3L_1(3L_3 - 1)$ $N_9 = \frac{9}{2}L_3L_1(3L_1 - 1), N_{10} = 27L_1L_2L_3$

2) 矩形单元 二维母单元是  $(\xi, \eta)$  平面中的  $2 \times 2$  正方形, 其中

$$-1 \leq \xi \leq +1, -1 \leq \eta \leq +1$$

坐标原点放在单元形心上, 单元边界是 4 条直线:  $\xi = \pm 1, \eta = \pm 1$ , 单元节点放在角点及各边等分点上。常用矩形平面单元的形函数见表 40.2-2。

表 40.2-2 常用矩形平面单元的形函数

单元名称	单元(母单元)形态	自由度	形函数
4 节点平面等参元		$u, v$	$N_i = \frac{1}{4}(1 + \xi_i)(1 + \eta_i)$ $(i=1, 2, 3, 4)$
8 节点平面等参元		$u, v$	$N_i = \frac{1}{4}(1 + \xi_i)(1 + \eta_i)(\xi_i + \eta_i - 1) \quad (i=1, 3, 5, 7)$ $N_i = \frac{1}{2}(1 - \xi^2)(1 - \eta_i) \quad (i=2, 6)$ $N_i = \frac{1}{2}(1 - \eta^2)(1 - \xi_i) \quad (i=4, 8)$
12 节点平面等参元		$u, v$	$N_i = \frac{1}{32}(1 + \xi_i)(1 + \eta_i)[-10 + 9(\xi^2 + \eta^2)] \quad (i=1, 4, 7, 10)$ $N_i = \frac{9}{32}(1 - \xi_i)(1 - \eta^2)(1 + 9\eta_i) \quad (i=5, 6, 11, 12)$ $N_i = \frac{9}{32}(1 + \eta_i)(1 - \xi^2)(1 + 9\xi_i) \quad (i=2, 3, 8, 9)$
备注	$\xi_0 = \xi, \xi, \eta_0 = \eta, \eta$		

通过坐标变换, 将正方形母单元变换成整体坐标系中任意四边形子单元。在整体坐标系中, 二维子单元内任意一点的坐标用形函数表示如下:

$$x = \sum N_i x_i = N_1 x_1 + N_2 x_2 + \dots$$

$$y = \sum N_i y_i = N_1 y_1 + N_2 y_2 + \dots$$

式中  $N_i$  —— 用自然坐标或面积坐标表示的形函数;

$(x_i, y_i)$  —— 节点  $i$  的整体坐标。

### 3 三维等参单元

1) 四面体单元 四面体单元的形函数往往用体积坐标来表示, 体积坐标如图 40.2-3 所示。常用的四面体单元的形函数见表 40.2-3。

2) 六面体单元 三维母单元是  $(\xi, \eta, \zeta)$  坐标系中的  $2 \times 2 \times 2$  正六面体, 其中

$$-1 \leq \xi \leq +1, -1 \leq \eta \leq +1, -1 \leq \zeta \leq +1$$

坐标原点在单元形心上,单元边界是 6 个平面; $\xi = \pm 1, \eta = \pm 1, \zeta = \pm 1$ ,单元结点放在角点及各棱边的等分点上。常用的六面体单元的形函数见表 40.2-3。

3) 三棱柱单元 三棱柱单元的形函数在平行于棱边的方向坐标  $\zeta$ ,在垂直于棱边的三角形平面内用面积坐标  $L_i$ ,常用的三棱柱单元的形函数见表 40.2-3。

通过坐标变换,将空间母单元变换成整体坐标系中的曲边曲面空间子单元。在整体坐标系中,空间子单元内任意一点的坐标用形函数表示如下:

$$\begin{aligned} x &= \sum N_i x_i = N_1 x_1 + N_2 x_2 + \dots \\ y &= \sum N_i y_i = N_1 y_1 + N_2 y_2 + \dots \\ z &= \sum N_i z_i = N_1 z_1 + N_2 z_2 + \dots \end{aligned}$$

式中  $N_i$  —— 用自然坐标或体积坐标及面积坐标表示的形函数;

$(x_i, y_i, z_i)$  —— 结点  $i$  的整体坐标。

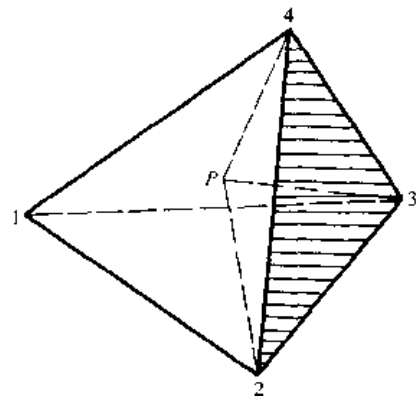
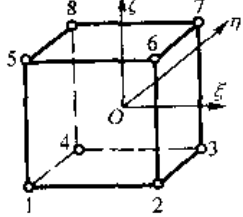
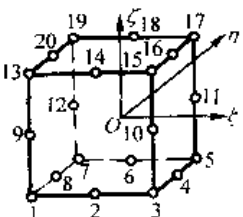
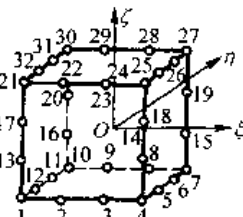
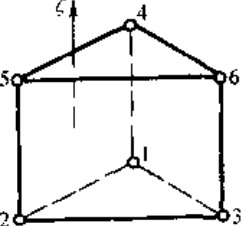
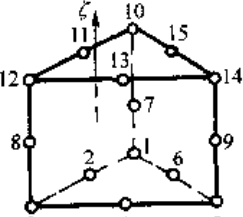


图 40.2-3 体积坐标

表 40.2-3 常用空间单元的形函数

单元名称	单元(母单元)形态	自由度	形函数
4 节点线性四面体单元		$u, v, w$	$\begin{aligned} N_1 &= L_1 \\ N_2 &= L_2 \\ N_3 &= L_3 \\ N_4 &= L_4 \end{aligned}$
10 节点二次四面体单元		$u, v, w$	$\begin{aligned} N_1 &= (2L_1 - 1)L_1, N_2 = 4L_1L_2 \\ N_3 &= (2L_2 - 1)L_2, N_4 = 4L_2L_3 \\ N_5 &= (2L_3 - 1)L_3, N_6 = 4L_3L_1 \\ N_7 &= 4L_1L_4, N_8 = 4L_2L_4 \\ N_9 &= 4L_3L_4, N_{10} = (2L_4 - 1)L_4 \end{aligned}$
20 节点三次四面体单元		$u, v, w$	$\begin{aligned} N_1 &= \frac{1}{2}(3L_1 - 1)(3L_1 - 2)L_1, N_2 = \frac{9}{2}L_1L_2(3L_1 - 1) \\ N_3 &= \frac{9}{2}L_1L_2(3L_2 - 1), N_4 = \frac{1}{2}(3L_2 - 1)(3L_2 - 2)L_2 \\ N_5 &= \frac{9}{2}L_2L_3(3L_2 - 1), N_6 = \frac{9}{2}L_3L_2(3L_3 - 1) \\ N_7 &= \frac{1}{2}(3L_3 - 1)(3L_3 - 2)L_3, N_8 = \frac{9}{2}L_3L_1(3L_3 - 1) \\ N_9 &= \frac{9}{2}L_3L_1(3L_1 - 1), N_{10} = 27L_1L_2L_3 \\ N_{11} &= \frac{9}{2}L_1L_4(3L_1 - 1), N_{12} = \frac{9}{2}L_2L_4(3L_2 - 1) \\ N_{13} &= \frac{9}{2}L_3L_4(3L_3 - 1), N_{14} = \frac{9}{2}L_1L_4(3L_4 - 1) \\ N_{15} &= \frac{9}{2}L_2L_4(3L_4 - 1), N_{16} = \frac{9}{2}L_3L_4(3L_4 - 1) \\ N_{17} &= \frac{1}{2}(3L_1 - 1)(3L_4 - 2)L_4, N_{18} = 27L_2L_3L_4 \\ N_{19} &= 27L_1L_3L_4, N_{20} = 27L_1L_2L_4 \end{aligned}$

(续)

单元名称	单元(母单元)形态	自由度	形函数
8 节点空间 等参元		$u, v, w$	$N_i = \frac{1}{8} (1 \pm \xi_0)(1 \pm \eta_0)(1 \pm \zeta_0)$ $(i=1, 2, \dots, 8)$
20 节点空间 等参元		$u, v, w$	$N_i = \frac{1}{8} (1 + \xi_0)(1 + \eta_0)(1 + \zeta_0)(\xi_0 + \eta_0 + \zeta_0 - 2)$ $(i=1, 3, 5, 7, 13, 15, 17, 19)$ $N_i = \frac{1}{4} (1 - \xi^2)(1 + \eta_0)(1 + \zeta_0) \quad (i=2, 6, 14, 18)$ $N_i = \frac{1}{4} (1 - \eta^2)(1 + \xi_0)(1 + \zeta_0) \quad (i=4, 8, 16, 20)$ $N_i = \frac{1}{4} (1 - \zeta^2)(1 + \xi_0)(1 + \eta_0) \quad (i=9, 10, 11, 12)$
32 节点空间 等参元		$u, v, w$	$N_i = \frac{1}{64} (1 + \xi_0)(1 + \eta_0)(1 + \zeta_0)[9(\xi^2 + \eta^2 + \zeta^2) - 19]$ $(i=1, 4, 7, 10, 21, 24, 27, 30)$ $N_i = \frac{9}{64} (1 - \xi^2)(1 + 9\xi_0)(1 + \eta_0)(1 + \zeta_0)$ $(i=2, 3, 8, 9, 22, 23, 28, 29)$ $N_i = \frac{9}{64} (1 - \eta^2)(1 - 9\eta_0)(1 + \xi_0)(1 + \zeta_0)$ $(i=5, 6, 11, 12, 25, 26, 31, 32)$ $N_i = \frac{9}{64} (1 - \zeta^2)(1 + 9\xi_0)(1 + \xi_0)(1 + \eta_0)$ $(i=13, 14, 15, 16, 17, 18, 19, 20)$
6 节点线性 三棱柱元		$u, v, w$	$N_i = \frac{1}{2} L_1(1 + \zeta_0) \quad (i=1, 4)$ $N_i = \frac{1}{2} L_2(1 + \zeta_0) \quad (i=2, 5)$ $N_i = \frac{1}{2} L_3(1 + \zeta_0) \quad (i=3, 6)$
15 节点二次 三棱柱元		$u, v, w$	$N_i = \frac{1}{2} L_1(2L_1 - 1)(1 - \zeta_0) - \frac{1}{2} L_1(1 - \zeta_0) \quad (i=1, 10)$ $N_i = \frac{1}{2} L_2(2L_2 - 1)(1 + \zeta_0) - \frac{1}{2} L_2(1 - \zeta_0) \quad (i=3, 12)$ $N_i = \frac{1}{2} L_3(2L_3 - 1)(1 + \zeta_0) - \frac{1}{2} L_3(1 - \zeta_0) \quad (i=5, 14)$ $N_i = 2L_3 L_2(1 + \zeta_0) \quad (i=2, 11)$ $N_i = 2L_2 L_3(1 + \zeta_0) \quad (i=4, 13)$ $N_i = 2L_3 L_1(1 + \zeta_0) \quad (i=6, 15)$ $N_7 = L_1(1 - \zeta^2), N_8 = L_2(1 - \zeta^2), N_9 = L_3(1 - \zeta^2)$
备注	$\xi_0 = \xi, \eta_0 = \eta, \zeta_0 = \zeta; L_i$ 在四面体单元的形函数中为体积坐标, 在三棱柱单元的形函数中为面积坐标。		

#### 4 等参数单元用于弹性力学分析的 一般格式

等参数单元通常也以位移作为基本未知量,与广义坐标有限元法的一般格式对等参元同样适用,由于等参元的形函数是使用自然坐标给出的,等参元的一切计算都是在自然坐标中规则的母单元内进行,因此需要作坐标转换对广义坐标有限元法的一般格式加以修正得到等参元的一般格式。

1) 母单元为  $\xi, \eta, \zeta$  坐标系中的立方体单元系列。

自然坐标有

$$-1 \leq \xi \leq +1, -1 \leq \eta \leq +1, -1 \leq \zeta \leq +1$$

单元矩阵计算时,单元刚度矩阵和节点载荷列阵的表达式变为如下形式:

单元刚度矩阵

$$K^e = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \mathbf{B}^T \mathbf{D} \mathbf{B} |J| d\xi d\eta d\zeta$$

分布体力的单元等效节点载荷

$$\{P\}_f^e = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \mathbf{N}^T \{f\} |J| d\xi d\eta d\zeta$$

分布面力的单元等效节点载荷

$$\{P\}_q^e = \int_{-1}^1 \int_{-1}^1 \mathbf{N}^T \{q\} |J| \Delta d\eta d\zeta$$

( $\{q\}$  作用在  $\xi = 1$  的面)

初应变与初应力的单元等效节点载荷

$$\{P\}_{\epsilon_0}^e = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \mathbf{B}^T \mathbf{D} \{\epsilon_0\} |J| d\xi d\eta d\zeta$$

$$\{P\}_{\sigma_0}^e = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \mathbf{B}^T \{\sigma_0\} |J| d\xi d\eta d\zeta$$

其中

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{vmatrix}$$

$$\Delta = \left[ \left( \frac{\partial y}{\partial \eta} \frac{\partial z}{\partial \xi} - \frac{\partial y}{\partial \xi} \frac{\partial z}{\partial \eta} \right)^2 + \left( \frac{\partial z}{\partial \eta} \frac{\partial x}{\partial \xi} - \frac{\partial z}{\partial \xi} \frac{\partial x}{\partial \eta} \right)^2 + \left( \frac{\partial x}{\partial \eta} \frac{\partial y}{\partial \xi} - \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} \right)^2 \right]^{\frac{1}{2}}$$

2) 母单元为体积坐标的四面体单元系列。自然坐标取体积坐标  $L_1, L_2, L_3, L_4$ , 且它们不完全独立,令

$$\xi = L_1, \eta = L_2, \zeta = L_3, 1 - \xi - \eta - \zeta = L_4$$

单元矩阵计算时,单元刚度矩阵和节点载荷列阵的表达式变为如下形式:

单元刚度矩阵

$$K^e = \int_0^1 \int_0^{1-L_1} \int_0^{1-L_1-L_2} \mathbf{B}^T \mathbf{D} \mathbf{B} |J| dL_1 dL_2 dL_3$$

分布体力的单元等效节点载荷

$$\{P\}_f^e = \int_0^1 \int_0^{1-L_1} \int_0^{1-L_1-L_2} \mathbf{N}^T \{f\} |J| dL_1 dL_2 dL_3$$

分布面力的单元等效节点载荷

$$\{P\}_q^e = \int_0^1 \int_0^{1-L_1} \mathbf{N}^T \{q\} |J| \Delta dL_2 dL_3$$

( $\{q\}$  作用在  $L_1 = 0$  的面)

初应力与初应变的单元等效节点载荷

$$\{P\}_{\epsilon_0}^e = \int_0^1 \int_0^{1-L_1} \int_0^{1-L_1-L_2} \mathbf{B}^T \mathbf{D} \{\epsilon_0\} |J| dL_1 dL_2 dL_3$$

$$\{P\}_{\sigma_0}^e = \int_0^1 \int_0^{1-L_1} \int_0^{1-L_1-L_2} \mathbf{B}^T \{\sigma_0\} |J| dL_1 dL_2 dL_3$$

其中

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial L_1} & \frac{\partial y}{\partial L_1} & \frac{\partial z}{\partial L_1} \\ \frac{\partial x}{\partial L_2} & \frac{\partial y}{\partial L_2} & \frac{\partial z}{\partial L_2} \\ \frac{\partial x}{\partial L_3} & \frac{\partial y}{\partial L_3} & \frac{\partial z}{\partial L_3} \end{vmatrix}$$

$$\Delta = \left[ \left( \frac{\partial y}{\partial L_2} \frac{\partial z}{\partial L_3} - \frac{\partial y}{\partial L_3} \frac{\partial z}{\partial L_2} \right)^2 + \left( \frac{\partial z}{\partial L_2} \frac{\partial x}{\partial L_3} - \frac{\partial z}{\partial L_3} \frac{\partial x}{\partial L_2} \right)^2 + \left( \frac{\partial x}{\partial L_2} \frac{\partial y}{\partial L_3} - \frac{\partial x}{\partial L_3} \frac{\partial y}{\partial L_2} \right)^2 \right]^{\frac{1}{2}}$$

对于二维问题和一维问题只需要将以上各公式退化就可以得到母单元为正方形系列和三角形系列的二维等参元以及直线系列的一维等参元的相应公式。

## 5 数值积分方法

### 5.1 一维数值积分

求解一维问题的数值积分  $\int_a^b F(\xi) d\xi$ , 积分点  $\xi_i$  的数目和位置决定了数值积分的精度。对于  $n$  个积分点, 按照积分点位置的不同选择, 通常采用两种不同的数值积分方案, 即 Newton-Cotes 积分方案和高斯积分方案。

1) Newton-Cotes 积分 Newton-Cotes 积分中, 积分点的位置按等间距分布, 即

$$\xi_i = a + ih \quad (i=0, 1, 2, \dots, n-1)$$

其中  $h$  是积分点间距,  $h = (b-a)/(n-1)$ 。

Newton-Cotes 数值积分式如下:

$$\int_a^b F(\xi) d\xi = \sum_{i=1}^n H_i F(\xi_i) + R_{n-1}$$

式中  $R_{n-1}$  是余项,  $H_i$  称为积分的加权系数, 简称权。

$$H_i = (b-a)C_i^{n-1}$$

其中

$$C_i^{n-1} = \int_0^1 t_i^{n-1}(\xi') d\xi'$$

$C_i^{n-1}$  称为  $n-1$  阶的 Newton-Cotes 数值积分常数。式中

$$t_i^{n-1}(\xi') = \frac{(\xi' - \xi'_1)(\xi' - \xi'_2)\cdots(\xi' - \xi'_{i-1})(\xi' - \xi'_{i+1})\cdots(\xi' - \xi'_n)}{(\xi'_i - \xi'_1)(\xi'_i - \xi'_2)\cdots(\xi'_i - \xi'_{i-1})(\xi'_i - \xi'_{i+1})\cdots(\xi'_i - \xi'_n)}$$

$$\xi' = \frac{1}{b-a}(\xi - a)$$

2) 一维高斯积分 在一维高斯积分中,积分点  $\xi_i$  不是等间距分布的,积分点  $\xi_i$  应是勒让德多项式  $L_n(\xi)$  的根。

一维高斯数值积分式如下:

$$\int_a^b F(\xi) d\xi = \sum_{i=1}^n H_i F(\xi_i) + R_{n-1}$$

其中加权系数  $H_i$  按下式计算:

$$H_i = \frac{2}{(1-\xi_i^2)[L'_n(\xi_i)]^2}$$

## 5.2 二维及三维高斯积分

对于二重积分

$$\int_{-1}^1 \int_{-1}^1 F(\xi, \eta) d\xi d\eta = \sum_{i=1}^n \sum_{j=1}^n H_i H_j F(\xi_i, \eta_j) + R_{n-1}$$

对于三重积分

$$\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 F(\xi, \eta, \zeta) d\xi d\eta d\zeta = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n H_i H_j H_k F(\xi_i, \eta_j, \zeta_k) + R_{n-1}$$

## 6 等参元计算中数值积分阶次的选择

当应用数值积分进行求解时,数值积分的阶次将直接影响计算的精度和工作量,如果选择不当,会导致计算数据的不准确,甚至计算失败。

选择积分阶次的原则如下:

- 1) 保证积分的精度。
- 2) 保证结构总刚度矩阵  $K$  是非奇异的。

### 第3章 杆件结构力学问题的有限单元法

杆件结构在机械工程中应用广泛，有不少机器的机架都是由杆件构成的。杆件结构分析通过确定结构的位移、应力和应变等，为机械设计中的强度、刚度和稳定性等分析计算提供依据。

形分析时，假定应力在截面上均匀分布，原来垂直于轴线的截面变形后仍保持和轴线垂直。轴力杆单元只承受轴向载荷，节点位移也只有轴向位移，其中2节点杆单元的坐标及位移如图40.3-1所示。轴力杆单元的分析公式见表40.3-1。

#### 1 等截面直杆—梁单元

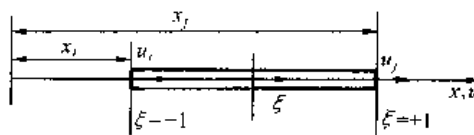


图 40.3-1 2节点杆单元

#### 1.1 轴力杆单元

承受轴向载荷的等截面直杆，对其进行应力和变

表 40.3-1 轴力杆单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [U_1 \ U_2 \ \dots \ U_n]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u_1 \ u_2 \ \dots \ u_n]$
3	位移函数	$u$	$u = [N_1 \ N_2 \ \dots \ N_n] \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{Bmatrix} = N\delta^e$ <p>如2节点单元 <math>N_1 = \frac{1}{2}(1-\xi)</math>, <math>N_2 = \frac{1}{2}(1+\xi)</math></p>
4	应变	$\epsilon_x^e$	$\epsilon_x^e = B\delta^e$ $B = \left[ \frac{dN_1}{dx} \ \frac{dN_2}{dx} \ \dots \ \frac{dN_n}{dx} \right]$
5	应力	$\sigma_x^e$	$\sigma_x^e = E\epsilon_x^e = EB\delta^e$
6	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \int_0^l EA \left( \frac{dN}{dx} \right)^T \left( \frac{dN}{dx} \right) dx = \int_{-1}^1 \frac{2EA}{l} \left( \frac{dN}{d\xi} \right)^T \left( \frac{dN}{d\xi} \right) d\xi$ <p>如2节点单元 <math>\bar{K}^e = \frac{EA}{l} \begin{bmatrix} 1 &amp; -1 \\ -1 &amp; 1 \end{bmatrix}</math></p>
7	等效节点载荷	$F^e$	$F^e = \int_0^l N^T f(x) dx = \int_{-1}^1 N^T f(\xi) \frac{l}{2} d\xi$
备注	$N_1, N_2, \dots, N_n$ —单元的形函数, 简称形函数; $N$ —形函数矩阵; $n$ —单元的节点数; $\xi$ —单元内的自然坐标; $B$ —几何矩阵; $l$ —杆单元的长度; $A$ —杆单元的截面积; $f(x), f(\xi)$ —分布载荷。		

#### 1.2 扭转杆单元

受扭矩作用的等截面直杆单元和受轴力作用的等截面直杆单元属于同一类型单元，求解扭转杆单元的方程和表达式只需要将求解轴力杆单元的相应的方程和表达式的变量和符号替换即可，见表40.3-2。并且

这些方程和表达式只适用于自由扭转情况，即扭转变形后，截面保持平面，不发生翘曲。

表 40.3-2

轴力杆单元	扭转杆单元
$F^e$ (轴力)	$M^e$ (扭矩)
$u$ (轴向位移)	$\theta$ (扭转角)
$\epsilon_x^e$ (轴向应变)	$\alpha$ (截面扭转率)
$E$ (弹性模量)	$G$ (切变模量)
$A$ (轴力杆截面积)	$J$ (截面扭转惯性矩)

#### 1.3 弯曲梁单元

等截面梁单元承受横向作用的分布载荷、横向集中载荷和弯矩作用。在经典的梁弯曲理论中，假设垂直梁中心线的截面，变形后仍保持为平面，且仍垂直于中心线。梁单元端部有横向位移和转角，其中2节点梁单元的内力和位移如图40.3-2所示。梁单元的分析公式见表40.3-3。

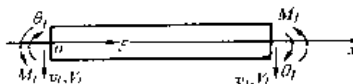


图 40.3-2 2节点梁单元



表 40.3-3 梁单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [V_i, M_i, V_j, M_j]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [v_i, \theta_i, v_j, \theta_j]^T$
3	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \frac{EI}{l^3} \begin{bmatrix} 12 & -6l & -12 & -6l \\ -6l & 4l^2 & 6l & 2l^2 \\ -12 & 6l & 12 & 6l \\ -6l & 2l^2 & 6l & 4l^2 \end{bmatrix}$
备注	$l$ —梁单元的长度; $I$ —梁单元的截面弯曲惯性矩。		

## 2 杆件系统结构分析

杆件系统是指由杆件组成的结构系统。平面杆件系统是指系统本身几何结构和所承受的载荷都处于同一平面内。空间杆件系统则不局限于同一平面内。杆件系统可以用杆—梁单元（以后简称为杆单元）进行离散。一般情况下，单元处于拉压、扭转和弯曲共同作用下，而不是某一种载荷单独作用。在进行结构分析时，首先建立结构共用的总体坐标系，然后通过坐标转换将各个建立于单元局部坐标系的单元特性矩阵转换到总体坐标系。

### 2.1 杆系结构的矩阵分析

杆系结构分析应用矩阵代数表达力学方程，可以使公式简洁、便利，具有统一性，而且适于用计算机进行计算。应用位移法分析杆件结构的解题步骤是：

1) 将杆件结构离散化成有限个单元，在杆系结构中通常以一根杆作为一个单元，不同截面或不同材料的杆件分属不同的单元，它们在杆端相互联接。

2) 计算每个单元的刚度矩阵，然后通过平衡条件建立整个杆件结构的总刚度矩阵。

3) 计算作用于每个节点（相邻单元交点）上的载荷向量。

4) 求解节点位移，并求出各杆的内力、应力、应变等。

### 2.2 平面杆系结构的分析公式

一等截面杆单元，它的两端节点分别用  $i$  和  $j$  表示，其局部坐标系  $\bar{xOy}$  中， $\bar{x}$  轴与杆单元的轴线重合， $\bar{y}$  和  $\bar{z}$  轴为截面的主惯性轴。不考虑杆的约束时，单元  $e$  共有 6 个杆端位移（线位移和转角），其分析公式见表 40.3-4。

表 40.3-4 平面杆系单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [U_i, V_i, M_i, U_j, V_j, M_j]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u_i, v_i, \theta_i, u_j, v_j, \theta_j]^T$
3	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \begin{bmatrix} \frac{EA}{l} & 0 & 0 & -\frac{EA}{l} & 0 & 0 \\ 0 & \frac{12EI}{l^3} & \frac{6EI}{l^2} & 0 & -\frac{12EI}{l^3} & \frac{6EI}{l^2} \\ 0 & \frac{6EI}{l^2} & \frac{4EI}{l} & 0 & -\frac{6EI}{l^2} & \frac{2EI}{l} \\ \frac{EA}{l} & 0 & 0 & -\frac{EA}{l} & 0 & 0 \\ 0 & -\frac{12EI}{l^3} & -\frac{6EI}{l^2} & 0 & \frac{12EI}{l^3} & -\frac{6EI}{l^2} \\ 0 & \frac{6EI}{l^2} & \frac{2EI}{l} & 0 & -\frac{6EI}{l^2} & \frac{4EI}{l} \end{bmatrix}$
4	坐标变换	$T$	$T = \begin{bmatrix} \alpha & \beta & 0 & 0 & 0 & 0 \\ -\beta & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & \beta & 0 \\ 0 & 0 & 0 & -\beta & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ 式中 $\alpha = \cos\varphi$ , $\beta = \sin\varphi$
备注	$l$ —梁单元的长度; $I$ —梁单元的截面弯曲惯性矩, $\varphi$ —局部坐标 $\bar{x}$ 轴和总体坐标 $x$ 轴之间的夹角。		

### 2.3 空间杆系结构

空间杆单元与平面杆单元的区别在于：杆件承受

轴力、弯矩和扭矩作用，而且弯矩可能同时作用在两个坐标面内。单元  $e$  的每个节点有 6 个自由度，即有 6 个广义位移和 6 个广义力。其分析公式见表 40.3-5。

表 40.3-5 空间杆件单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [U, V, M, M_x, M_y, M_z, U, V, W, M_x, M_y, M_z]^T$
2	节点位移矢量	$\delta^e$	$\delta^e = [u, v, w, \theta_x, \theta_y, \theta_z, u, v, w, \theta_x, \theta_y, \theta_z]^T$
3	单元刚度矩阵	$K^e$	$K^e = \begin{bmatrix} \frac{EA}{l} & 0 & 0 & 0 & 0 & 0 & -\frac{EA}{l} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{12EI_z}{l^3} & 0 & 0 & 0 & \frac{6EI_z}{l^2} & 0 & -\frac{12EI_z}{l^3} & 0 & 0 & 0 & \frac{6EI_z}{l^2} \\ 0 & 0 & \frac{12EI_y}{l^3} & 0 & -\frac{6EI_y}{l^2} & 0 & 0 & 0 & -\frac{12EI_y}{l^3} & 0 & -\frac{6EI_y}{l^2} & 0 \\ 0 & 0 & 0 & \frac{GJ}{l} & 0 & 0 & 0 & 0 & 0 & -\frac{GJ}{l} & 0 & 0 \\ 0 & 0 & -\frac{6EI_y}{l^2} & 0 & \frac{4EI_y}{l} & 0 & 0 & 0 & \frac{6EI_y}{l^2} & 0 & \frac{2EI_y}{l} & 0 \\ 0 & \frac{6EI_z}{l^2} & 0 & 0 & 0 & 4EI_z & 0 & -\frac{6EI_z}{l^2} & 0 & 0 & 0 & \frac{2EI_z}{l} \\ \frac{EA}{l} & 0 & 0 & 0 & 0 & 0 & \frac{EA}{l} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{12EI_x}{l^3} & 0 & 0 & 0 & -\frac{6EI_x}{l^2} & 0 & \frac{12EI_x}{l^3} & 0 & 0 & 0 & -\frac{6EI_x}{l^2} \\ 0 & 0 & -\frac{12EI_x}{l^3} & 0 & \frac{6EI_x}{l^2} & 0 & 0 & 0 & \frac{12EI_x}{l^3} & 0 & \frac{6EI_x}{l^2} & 0 \\ 0 & 0 & 0 & \frac{GJ}{l} & 0 & 0 & 0 & 0 & 0 & \frac{GJ}{l} & 0 & 0 \\ 0 & 0 & -\frac{6EI_x}{l^2} & 0 & \frac{2EI_x}{l} & 0 & 0 & 0 & \frac{6EI_x}{l^2} & 0 & \frac{4EI_x}{l} & 0 \\ 0 & \frac{6EI_x}{l^2} & 0 & 0 & 0 & \frac{2EI_x}{l} & 0 & -\frac{6EI_x}{l^2} & 0 & 0 & 0 & \frac{4EI_x}{l} \end{bmatrix}$
4	坐标变换	$T$	$T = \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \quad \lambda = \begin{bmatrix} \cos(x, \bar{x}) & \cos(x, \bar{y}) & \cos(x, \bar{z}) \\ \cos(y, \bar{x}) & \cos(y, \bar{y}) & \cos(y, \bar{z}) \\ \cos(z, \bar{x}) & \cos(z, \bar{y}) & \cos(z, \bar{z}) \end{bmatrix}$
备注	$l$ —单元的长度; $A$ —单元的截面面积; $I$ —单元的截面弯曲惯性矩; $J$ —单元的截面扭转惯性矩。		

## 第 4 章 平板弯曲问题的有限单元法

在受到垂直于板面的载荷后,薄板产生弯曲变形。如果板的厚度比其他两个方向的尺寸小得多,而且挠度比厚度又小得多,在分析平板弯曲问题时,可采用下列假设:

(1) 略平板厚度方向的正应力,并假定薄板的厚度没有变化;

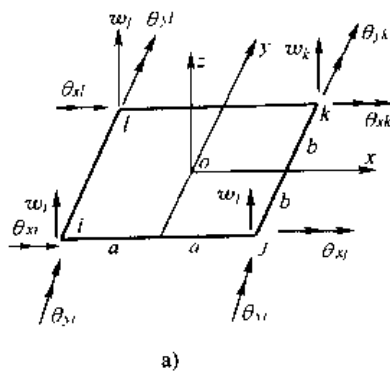
(2) 板中面内的各点没有平行于中面的位移;

(3) 薄板中面的法线在变形后仍保持法线。

利用上述假设将平板弯曲问题简化为二维问题,且全部应力和应变可用板中面的挠度  $w$  表示,中面的设定如图 40.4-1 所示。

### 1 非协调板单元

#### 1.1 矩形板单元



如图 40.4-2 所示,矩形板单元的 4 条边分别平行于  $x$  轴和  $y$  轴,每个节点 3 个自由度:挠度  $w$ 、绕  $x$  轴的转角  $\theta_x$  和绕  $y$  轴的转角  $\theta_y$ ,其分析公式见表 40.4-1。

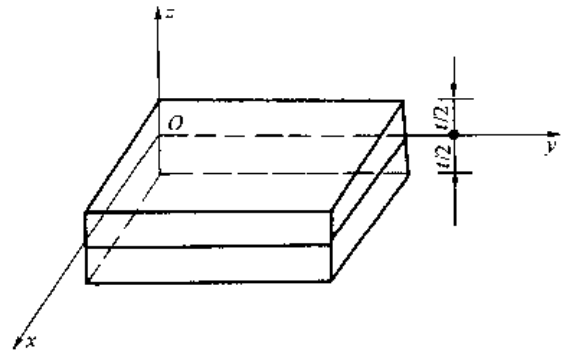


图 40.4-1 薄板中面

图 40.4-2 矩形板单元的节点位移与节点力

表 40.4-1 矩形板单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [F_i^x \ F_j^x \ F_k^x \ F_l^x]^T$ 式中 $F_i^x = [W_i \ M_{\theta_{xi}} \ M_{\theta_{yi}}]^T \ (i, j, k, l)$
2	节点位移矢量	$\delta^e$	$\delta^e = [\delta_i^x \ \delta_j^x \ \delta_k^x \ \delta_l^x]^T$ 式中 $\delta_i^x = [w_i \ \theta_{xi} \ \theta_{yi}]^T \ (i, j, k, l)$
3	位移函数	$w$	$w = \phi \beta = N \delta^e$ 式中 $\phi = [1 \ x \ y \ x^2 \ xy \ y^2 \ x^3 \ x^2y \ xy^2 \ y^3 \ x^3y \ xy^3]$ $\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_{12}]$ $N = [N_1 \ N_2 \ N_3 \ N_4]$ $N_i = \frac{1}{8} [(\xi_0+1)(\eta_0-1)(2-\xi_0+\eta_0-\xi_0^2-\eta_0^2) \ b\eta_0(\xi_0+1)(\eta_0+1)^2(\eta_0-1) - a\xi_0(\xi_0+1)^2(\xi_0-1)(\eta_0+1)]$

(续)

序号	类别	符号	公式
4	形变	$\psi$	$\psi = \left[ -\frac{\partial^2 w}{\partial x^2} \quad -\frac{\partial^2 w}{\partial y^2} \quad -2\frac{\partial^2 w}{\partial x \partial y} \right]^T - B\delta^e$ $B = - \begin{bmatrix} \frac{\partial^2 N_1}{\partial x^2} & \frac{\partial^2 N_{1r}}{\partial x^2} & \frac{\partial^2 N_{1s}}{\partial x^2} & \dots & \frac{\partial^2 N_{1t}}{\partial x^2} \\ \frac{\partial^2 N_1}{\partial y^2} & \frac{\partial^2 N_{1r}}{\partial y^2} & \frac{\partial^2 N_{1s}}{\partial y^2} & \dots & \frac{\partial^2 N_{1t}}{\partial y^2} \\ 2\frac{\partial^2 N_1}{\partial x \partial y} & 2\frac{\partial^2 N_{1r}}{\partial x \partial y} & 2\frac{\partial^2 N_{1s}}{\partial x \partial y} & \dots & 2\frac{\partial^2 N_{1t}}{\partial x \partial y} \end{bmatrix}$
5	应变	$\epsilon^e$	$\epsilon^e = z\psi$
6	内力	$M$	$M = D\psi = DB\delta^e$ $D = \frac{Et^3}{12(1-\mu^2)} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1-\mu}{2} \end{bmatrix}$
7	应力	$\sigma^e$	$\sigma^e = \frac{12}{t^3} zM$
8	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \iint B^T D B d_x d_y$
9	节点等效载荷		薄板表面法向均布载荷 $q$ $P_i^e = 4qab \left[ \frac{1}{4} \quad \frac{b}{12} \quad -\frac{a}{12} \quad \frac{1}{4} \quad \frac{b}{12} \quad \frac{a}{12} \quad \frac{1}{4} \quad -\frac{b}{12} \quad \frac{a}{12} \quad \frac{1}{4} \quad -\frac{b}{12} \quad -\frac{a}{12} \quad \dots \dots \right]^T$
备注	$a, b$ 矩形板单元的边长; $t$ —矩形板单元的厚度; $\xi, \eta$ —矩形板单元的局部坐标, 以矩形中心为原点, $\xi_0 = \xi/\xi_1, \eta_0 = \eta/\eta_1$		

### 1.2 三角形板单元

三角形板单元可以较好地适应复杂的边界形状, 实用价值较大, 采用较多的是 3 节点三角形板单元, 如

图 40.4-3 所示, 每个节点有 3 个自由度: 挠度  $w$ 、绕  $x$  轴的转角  $\theta_x$  和绕  $y$  轴的转角  $\theta_y$ , 其分析公式, 除位移函数外, 其余与矩形板单元的分析公式类似, 只是少一个节点。

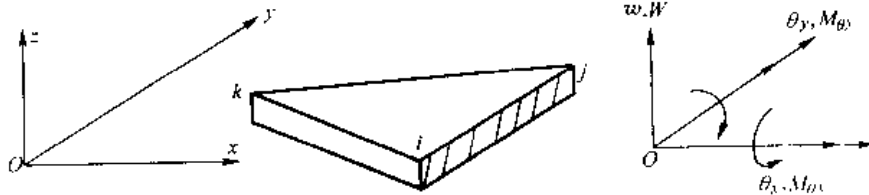


图 40.4-3 三角形板单元及其节点位移和节点力

用面积坐标表示的三角形板单元的位移函数为:

$$w = \beta_1 L_1 + \beta_2 L_2 + \beta_3 L_3$$

$$+ \beta_4 \left( L_2^2 L_1 + \frac{1}{2} L_1 L_2 L_3 \right)$$

$$+ \beta_5 \left( L_3^2 L_2 + \frac{1}{2} L_1 L_2 L_3 \right)$$

$$+ \beta_6 \left( L_1^2 L_3 - \frac{1}{2} L_1 L_2 L_3 \right)$$

$$+ \beta_7 \left( L_2^2 L_3 + \frac{1}{2} L_1 L_2 L_3 \right)$$

$$+ \beta_8 \left( L_1^2 L_3 + \frac{1}{2} L_1 L_2 L_3 \right)$$

$$+ \beta_9 \left( L_3^2 L_1 + \frac{1}{2} L_1 L_2 L_3 \right)$$

### 2 协调板单元

协调板单元在保持三个节点参数的条件下单元协调性的要求可以完全满足, 因而保证了有限元解的收敛性, 即在单元尺寸不断减小时, 解能单调收敛于精确解。但在实际计算中单元尺寸总是有限的, 因此计算结果常常使结构表现得过于刚硬。

在经典薄板理论的范围内, 使板单元满足协调性的要求的方法有两种, 一是增加节点参数, 即在节点参

数中还包括  $w$  的二次导数项；二是在保持每个节点有三个参数的前提下采取其他一些措施，如附加校正函数法、再分割法等。

以 3 节点三角形板单元为例，前述的 3 节点三角形板单元之所以是非协调的，是由于在每个边界上法向导数  $\partial w/\partial n$  是二次变化的，由两端节点不能唯一确定  $\partial w/\partial n$ 。因此需要增加与各个边界相联系的校正函数，例如和边 23 相联系的校正函数  $\phi_{23}$ ，它具有如下性质：

- 1) 在全部边界上  $\phi_{23}=0$ ；
- 2) 在边界 12 和 13 上法向导数  $\partial\phi_{23}/\partial n=0$ ；
- 3) 在边界 23 上  $\partial\phi_{23}/\partial n \neq 0$ ，按二次变化；并在边界 23 的中点 4 取单位值。

类似地可以有校正函数  $\phi_{31}$  和  $\phi_{12}$ ，然后按一定比例将这些校正函数叠加到原来的非协调位移函数中，结果如下：

$$w = w_0 + \gamma_1\phi_{23} + \gamma_2\phi_{31} + \gamma_3\phi_{12}$$

式中  $w_0$ ——非协调元的位移函数；

$\gamma_1, \gamma_2, \gamma_3$ ——待定常数。

通过调整它们的大小，使得  $\partial w/\partial n$  在各个边界中点的法向导数  $\partial w/\partial n$  成线性变化，因此两端的法向导

数值就能唯一确定  $\partial w/\partial n$  了，从而使相邻单元交界面上的协调性得以实现。校正函数可按式求得：

$$\phi_{23} = \frac{\epsilon_{23}}{(\partial\epsilon_{23}/\partial n)_4}$$

其中

$$\epsilon_{23} = \frac{L_1 L_2^2 L_3^2}{(L_1 + L_2)(L_1 + L_3)}$$

或

$$\epsilon_{23} = \frac{L_1 L_2^2 L_3^2 (1 + L_1)}{(L_1 + L_2)(L_1 + L_3)}$$

式中  $L_1, L_2, L_3$ ——面积坐标；

$(\partial\epsilon_{23}/\partial n)_4$ —— $\epsilon_{23}$  在边界 23 中点 4 的法向导数值。

类似地，可以得到另外两个校正函数  $\phi_{31}$  和  $\phi_{12}$ 。

### 3 Mindlin 板单元

按照 Mindlin 弹性板理论，计算中采用以下假设：

- 1) 板的挠度  $w$  是小的；
- 2) 变形前板中面的法线在变形后仍为直线，但不必垂直于变形后的中面；
- 3) 垂直于中面的正应力可以忽略。

根据上述假设，可得 Mindlin 板单元的分析公式，见表 40.4-2。

表 40.4-2 Mindlin 板单元的分析公式

序号	类别	符号	公式
1	节点力矢量	$F^e$	$F^e = [F_1^e \ F_2^e \ \dots \ F_s^e]^T$ 式中 $F_i^e = [W_i \ M_{\theta_{xi}} \ M_{\theta_{yi}}]^T \quad (i=1, 2, \dots, s)$
2	节点位移矢量	$\delta^e$	$\delta^e = [\delta_1^e \ \delta_2^e \ \dots \ \delta_s^e]^T$ 式中 $\delta_i^e = [w_i \ \theta_{xi} \ \theta_{yi}]^T \quad (i=1, 2, \dots, s)$
3	位移函数	$u$ $v$ $w$	$u = z\phi_y(x, y)$ $v = -z\phi_x(x, y)$ $w = w(x, y)$
4	应变	$\epsilon$ $\gamma$	$\epsilon = \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix} = z\psi \quad \gamma = \begin{Bmatrix} \gamma_{xz} \\ \gamma_{yz} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial w}{\partial y} - \phi_x \\ \frac{\partial w}{\partial x} - \phi_y \end{Bmatrix} = B_1 \delta^e$ 式中 $\psi = \left[ \frac{\partial\phi_y}{\partial x}, -\frac{\partial\phi_x}{\partial y}, \frac{\partial\phi_y}{\partial y}, \frac{\partial\phi_x}{\partial x} \right]^T = B_2 \delta^e$ 其中 $B_1 = \begin{bmatrix} 0 & 0 & \frac{\partial N_2}{\partial x} & \dots & 0 & 0 & \frac{\partial N_s}{\partial x} \\ 0 & -\frac{\partial N_1}{\partial y} & 0 & \dots & 0 & -\frac{\partial N_1}{\partial y} & 0 \\ 0 & \frac{\partial N_1}{\partial x} & \frac{\partial N_1}{\partial y} & \dots & 0 & -\frac{\partial N_1}{\partial x} & \frac{\partial N_s}{\partial y} \end{bmatrix}$ $B_2 = \begin{bmatrix} \frac{\partial N_1}{\partial y} & -N_1 & 0 & \dots & \frac{\partial N_1}{\partial y} & -N_s & 0 \\ \frac{\partial N_1}{\partial x} & 0 & N_1 & \dots & \frac{\partial N_1}{\partial x} & 0 & N_s \end{bmatrix}$

(续)

序号	类别	符号	公式
5	内力	$M$ $Q$	$M = D_b \psi, Q = D_s \gamma$ $D_b = \frac{Et^3}{12(1-\mu^2)} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1-\mu}{2} \end{bmatrix}, D_s = \frac{kEt}{2(1+\mu)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
6	应力	$\sigma$ $\tau$	$\sigma = \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = z \frac{E}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1-\mu}{2} \end{bmatrix} \psi$ $\tau = \begin{Bmatrix} \tau_{yz} \\ \tau_{zx} \end{Bmatrix} = \frac{kEt}{2(1+\mu)} \gamma$
8	单元刚度矩阵	$\bar{K}^e$	$\bar{K}^e = \int_{-1}^1 \int_{-1}^1 (B^T D_b B_s + B^T D_s B_s)  J  d\xi d\eta$
9	节点等效载荷		薄板表面法向均布载荷 $q$ $F^e = \int_{-1}^1 \int_{-1}^1 N^T q  J  d\xi d\eta$ 式中 $N = [N_1 \ 0 \ 0 \ N_2 \ 0 \ 0 \ \dots \ N_i \ 0 \ 0]$
备注	s—单元节点数; $t$ —矩形板单元的厚度; $k$ —考虑剪应力分布不均匀的系数; $N_i$ —矩形板单元的形状函数; $\xi, \eta$ —矩形单元的局部坐标, 原点在矩形的形心。		

#### 4 应力杂交板单元

基于修正余能原理的应力杂交板单元用于薄板弯曲问题, 并用矩阵表示, 可以写成

$$\Pi_{m_e} = \sum_r \left\{ \int_{\Omega_e} \frac{1}{2} M^T D^{-1} M d\Omega - \int_{\sigma_e} F^T \delta dS + \int_{\alpha_e} T^T \delta dS \right\}$$

式中  $M = [M_x \ M_y \ M_{xy}]^T$

$$D^{-1} = \frac{12}{Et^3} \begin{bmatrix} 1 & -\mu & 0 \\ -\mu & 1 & 0 \\ 0 & 0 & 2(1+\mu) \end{bmatrix}$$

$$T = \begin{Bmatrix} V_n \\ M_n \end{Bmatrix}, \delta = \begin{Bmatrix} w \\ -\frac{\partial w}{\partial n} \end{Bmatrix}, \bar{T} = \begin{Bmatrix} \bar{V}_n \\ \bar{M}_n \end{Bmatrix}$$

$$V_n = Q_n + \frac{\partial M_n}{\partial s}$$

式中  $\partial\Omega_e = S_e + S_{\sigma_e} + S_{\alpha_e}$

是单元  $\Omega_e$  的全部边界,  $S_e$  是单元  $e$  和其他单元相邻的边界面,  $S_{\sigma_e}$  是单元  $e$  的力的边界条件,  $S_{\alpha_e}$  是单元  $e$  的位移边界条件,  $t$  是板单元的厚度,  $n$  是法向方向,  $s$  是切向方向,  $T$  是弹性体边界上单位面积的内力,  $\bar{T}$  是弹性体单位面积上作用的面积力。

## 第 5 章 轴对称壳体问题的有限单元法

### 1 基于薄壳理论的轴对称壳体单元

#### 1.1 轴对称薄壳理论的基本公式

轴对称壳体中面上任一点位置由角  $\phi$  和  $\theta$  确定，

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \\ \gamma_{\theta\phi} \\ \kappa_s \\ \kappa_\theta \\ \kappa_{\theta\phi} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial u}{\partial s} + \frac{w}{R_s} \\ \frac{1}{r} \left( \frac{\partial v}{\partial \theta} + u \sin \phi + w \cos \phi \right) \\ \frac{\partial v}{\partial s} + \frac{1}{r} \left( \frac{\partial u}{\partial \theta} - v \sin \phi \right) \\ -\frac{\partial}{\partial s} \left( \frac{\partial w}{\partial s} - \frac{u}{R_s} \right) \\ -\frac{1}{r^2} \frac{\partial^2 w}{\partial \theta^2} + \frac{\cos \phi}{r^2} \frac{\partial v}{\partial \theta} - \frac{\sin \phi}{r} \left( \frac{\partial w}{\partial s} - \frac{u}{R_s} \right) \\ 2 \left( -\frac{1}{r} \frac{\partial^2 w}{\partial s \partial \theta} + \frac{\sin \phi}{r^2} \frac{\partial w}{\partial \theta} + \frac{\cos \phi}{r} \frac{\partial v}{\partial s} - \frac{\sin \phi \cos \phi}{r^2} v + \frac{1}{r R_s} \frac{\partial u}{\partial \theta} \right) \end{Bmatrix}$$

其中  $s$  表示经向的弧长,  $R_s$  是经向的曲率半径,  $r$  是平行圆半径(即中面上任一点的径向坐标)。  $\varepsilon_r, \varepsilon_\theta, \gamma_{\theta\phi}$  表示中面内的伸长和剪切,  $\kappa_s, \kappa_\theta, \kappa_{\theta\phi}$  表示中面曲率和扭率的变化。

已知中面的六个广义应变分量后, 中面外任一点的应变可以表示为

$$\varepsilon_r^{(z)} = \varepsilon_r + z \kappa_s, \varepsilon_\theta^{(z)} = \varepsilon_\theta + z \kappa_\theta, \gamma_{\theta\phi}^{(z)} = \gamma_{\theta\phi}$$

式中  $z$ ——该点至中面的距离(沿法线方向测量)。

广义应力分量与广义应变分量之间的弹性关系是

$$\boldsymbol{\sigma} = \begin{Bmatrix} N_s \\ N_\theta \\ N_{\theta\phi} \\ M_s \\ M_\theta \\ M_{\theta\phi} \end{Bmatrix} = \frac{Et}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 & 0 & 0 & 0 \\ \mu & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\mu}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{t^2}{12} & \frac{t^2 \mu}{12} & 0 \\ 0 & 0 & 0 & \frac{t^2 \mu}{12} & \frac{t^2}{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{t^2(1-\mu)}{24} \end{bmatrix} \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \\ \gamma_{\theta\phi} \\ \kappa_s \\ \kappa_\theta \\ \kappa_{\theta\phi} \end{Bmatrix}$$

位移可由其经(子午)向分量  $u$ , 周向分量  $v$  和法向分量  $w$  确定, 如图 41.5-1 所示。在薄壳理论中, 壳体内任一点的应变根据 Kirchhoff 直法线假设, 可通过中面的六个广义应变分量来描述, 它们和中面位移的关系是

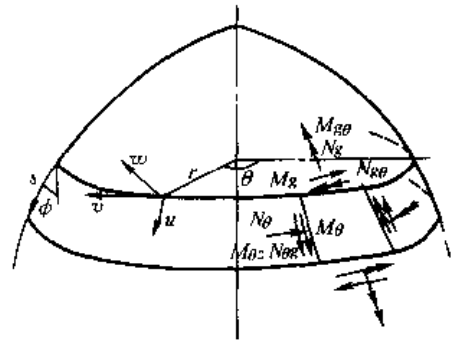


图 40.5-1 轴对称壳的坐标、位移和内力

式中  $N_s, N_\theta, N_{s\theta}$ ——壳体内垂直于  $s$  或  $\theta$  方向的截面上单位长度的内力;

$M_s, M_\theta, M_{s\theta}$  是相应截面上单位长度的力矩。

壳体内任一点的应力按下式计算

$$\sigma_s = \frac{N_s}{t} + \frac{12M_s}{t^3}z, \sigma_\theta = \frac{N_\theta}{t} + \frac{12M_\theta}{t^3}z,$$

$$\tau_{s\theta} = \frac{N_{s\theta}}{t} + \frac{12M_{s\theta}}{t^3}z$$

式中  $t$ ——壳体的厚度。

### 1.2 薄壳截锥单元

薄壳截锥单元是沿子午线方向将轴对称壳体划分为若干个线性单元,单元形状如图 40.5-2 所示。在每一单元内,角  $\phi$  是不变的。根据 Kirchhoff 直法线假设,在轴对称壳体所承受的载荷以及壳体的支承条件都是对称的,相应的应变和内力表达式蜕化为

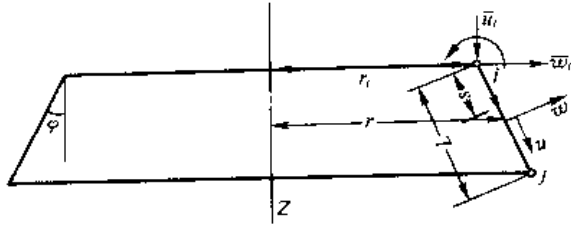


图 40.5-2 轴对称壳体的截锥单元

$$\epsilon = \begin{Bmatrix} \epsilon_s \\ \epsilon_\theta \\ \kappa_s \\ \kappa_\theta \end{Bmatrix} = \begin{Bmatrix} \frac{du}{ds} \\ \frac{1}{r}(u \sin\phi + w \cos\phi) \\ -\frac{d}{ds} \left( \frac{dw}{ds} - \frac{u}{R_i} \right) \\ -\frac{\sin\phi}{r} \left( \frac{dw}{ds} - \frac{u}{R_i} \right) \end{Bmatrix}$$

$$\sigma = \begin{Bmatrix} N_s \\ N_\theta \\ M_s \\ M_\theta \end{Bmatrix}$$

$$= \frac{Et}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 & 0 \\ \mu & 1 & 0 & 0 \\ 0 & 0 & \frac{t^2}{12} & \frac{t^2\mu}{12} \\ 0 & 0 & \frac{t^2\mu}{12} & \frac{t^2}{12} \end{bmatrix} \begin{Bmatrix} \epsilon_s \\ \epsilon_\theta \\ \kappa_s \\ \kappa_\theta \end{Bmatrix}$$

每个单元有 2 个节点,节点编号分别为  $i, j$ ,在轴对称载荷作用下,每个单元的位移参数为

$$\delta^e = \begin{Bmatrix} \delta_i \\ \delta_j \end{Bmatrix} = [\bar{u}_i \quad \bar{w}_i \quad \beta_i \quad \bar{u}_j \quad \bar{w}_j \quad \beta_j]^T$$

式中  $\bar{u}_i, \bar{w}_i$ ——整体坐标系中节点  $i$  的轴向和径向位移分量;

$\beta_i$ ——经向切线的转角。

局部坐标系中的单元节点位移与整体坐标下的节点位移存在如下关系:

$$\begin{Bmatrix} u_i \\ w_i \\ (dw/ds)_i \end{Bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} \bar{u}_i \\ \bar{w}_i \\ \beta_i \end{Bmatrix}$$

$$= T \delta^e \quad (i, j)$$

局部坐标系中单元的位移函数为

$$\begin{Bmatrix} u \\ w \end{Bmatrix} = [N_s, N_j] \begin{bmatrix} u_i & w_i & (dw/ds)_i & u_j & w_j & (dw/ds)_j \end{bmatrix}^T$$

$$N_i = \begin{bmatrix} 1-\xi & 0 & 0 \\ 0 & 1-3\xi^2-2\xi^3 & L(\xi-2\xi^2+\xi^3) \end{bmatrix}$$

其中

$$N_j = \begin{bmatrix} \xi & 0 & 0 \\ 0 & 3\xi^2-2\xi^3 & L(-\xi^2+\xi^3) \end{bmatrix}$$

式中  $\xi = s/L$ ;

$L$ ——薄壳截锥单元经线的长度。

在整体坐标系中单元的位移函数为

$$\begin{Bmatrix} u \\ w \end{Bmatrix} = [N_i T \quad N_j T] \begin{Bmatrix} \delta_i \\ \delta_j \end{Bmatrix} = N \delta^e$$

薄壳截锥单元的应变为

$$\epsilon^e = B \delta^e = [B_i T \quad B_j T] \delta^e$$

$$B_i = \begin{bmatrix} -1/L & 0 & 0 \\ (1-\xi)\sin\phi/r & (1-3\xi^2+2\xi^3)\cos\phi/r & L(1-2\xi^2+\xi^3)\cos\phi/r \\ 0 & 6(1-2\xi)/L^2 & (4-6\xi)/L \\ 0 & 6\xi(1-\xi)\sin\phi/(rL) & (-1+4\xi-3\xi^2)\sin\phi/r \end{bmatrix}$$

$$B_j = \begin{bmatrix} 1/L & 0 & 0 \\ \xi\sin\phi/r & (3\xi^2-2\xi^3)\cos\phi/r & L(-\xi^2+\xi^3)\cos\phi/r \\ 0 & -6(1-2\xi)/L^2 & 2(1-3\xi)/L \\ 0 & -6\xi(1-\xi)\sin\phi/(rL) & (2\xi-3\xi^2)\sin\phi/r \end{bmatrix}$$

薄壳截锥单元的刚度矩阵由下式计算:



$$\bar{K}^e = \int_0^1 2\pi r L B^T D B d\xi = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

式中  $K_{ij} = 2\pi L T^T \left( \int_0^1 B_i^T D B_j r d\xi \right) T$

侧向分布载荷  $p = [p_n \quad p_w]^T$ , 其中  $p_n, p_w$  分别为经向和法向的载荷分量, 截锥单元的等效节点载荷向量为

$$P^e = 2\pi L T^T \int_0^1 N^T p r d\xi \quad (i, j)$$

且  $P_i^e = [P_{z_i} \quad P_{w_i} \quad P_{\beta_i}]^T$

其中  $P_{z_i}, P_{w_i}, P_{\beta_i}$  分别为沿  $z, r$  方向的力和沿  $\beta$  方向的力矩。

## 2 位移和转动各自独立插值的轴对称壳体单元

### 2.1 考虑横向剪切变形的轴对称壳体理论的基本公式

考虑横向剪切的轴对称壳, 在受轴对称载荷作用下, 中面广义应变和位移的关系式如下:

$$\epsilon = \begin{Bmatrix} \epsilon_r \\ \epsilon_\theta \\ \kappa_r \\ \kappa_\theta \\ \gamma \end{Bmatrix} = \begin{Bmatrix} \frac{du}{ds} + \frac{w}{R} \\ \frac{1}{r}(u \sin\phi + w \cos\phi) \\ -\frac{d\beta}{ds} \\ -\frac{\sin\phi}{r}\beta \\ \frac{dw}{ds} - \frac{u}{R} - \beta \end{Bmatrix}$$

式中  $\epsilon_r, \epsilon_\theta$  —— 中面的经向应变和环向应变;

$\kappa_r, \kappa_\theta$  —— 中面的经向曲率变化和环向曲率变化;

$\gamma$  —— 横向切应变。

相应的中面内力(广义应力)

$$\sigma = [N_r \quad N_\theta \quad M_r \quad M_\theta \quad V]^T$$

依次是经向内力、环向内力、经向弯矩、环向弯矩和横向剪力。

广义应力和应变之间的关系是

$$\sigma = D \epsilon$$

其中  $D$  是弹性矩阵

$$D = \begin{bmatrix} D^m & 0 & 0 \\ 0 & D^b & 0 \\ 0 & 0 & D^s \end{bmatrix}$$

式中  $D^m, D^b, D^s$  —— 薄膜、弯曲和剪切刚度。

对于各向同性材料

$$D^m = \frac{Et}{1-\mu^2} \begin{bmatrix} 1 & \mu \\ \mu & 1 \end{bmatrix}, D^b = \frac{t^3}{12} D^m$$

$$D^s = k \frac{Et}{2(1+\mu)}, k = \frac{5}{6}$$

壳体应变能表达式是

$$\begin{aligned} U &= \frac{1}{2} \int_\Omega \epsilon^T D \epsilon d\Omega \\ &= \frac{1}{2} \int_\Omega (\epsilon^m)^T D^m \epsilon^m d\Omega \\ &\quad + \frac{1}{2} \int_\Omega (\epsilon^b)^T D^b \epsilon^b d\Omega \\ &\quad + \frac{1}{2} \int_\Omega \gamma D^s \gamma d\Omega \end{aligned}$$

其中

$$\epsilon^m = \begin{bmatrix} \epsilon_r \\ \epsilon_\theta \end{bmatrix}, \epsilon^b = \begin{bmatrix} \kappa_r \\ \kappa_\theta \end{bmatrix}$$

### 2.2 截锥单元

截锥单元较薄壳截锥单元增加了截面转动  $\beta$ , 它是独立的函数。对于两节点的截锥单元, 位移函数如下:

$$\bar{u} = \sum_{i=1}^2 N_i \bar{u}_i,$$

$$\bar{w} = \sum_{i=1}^2 N_i \bar{w}_i, \beta = \sum_{i=1}^2 N_i \beta_i,$$

其中  $N_1 = 1 - \xi, N_2 = \xi, \xi = s/L$

用总体坐标表示的应变表达式

$$\epsilon = \begin{Bmatrix} \epsilon_r \\ \epsilon_\theta \\ \kappa_r \\ \kappa_\theta \\ \gamma \end{Bmatrix} = \begin{bmatrix} \cos\phi \frac{d}{ds} & \sin\phi \frac{d}{ds} & 0 \\ 0 & \frac{1}{r} & 0 \\ 0 & 0 & -\frac{d}{ds} \\ 0 & 0 & -\frac{\sin\phi}{r} \\ -\sin\phi \frac{d}{ds} & \cos\phi \frac{d}{ds} & -1 \end{bmatrix} \times$$

$$\begin{Bmatrix} \bar{u} \\ \bar{w} \\ \beta \end{Bmatrix} = [B_1 \quad B_2] \begin{Bmatrix} \delta_1 \\ \delta_2 \end{Bmatrix}$$

其中

$$B_i = \begin{bmatrix} \cos\phi \frac{dN_i}{ds} & \sin\phi \frac{dN_i}{ds} & 0 \\ 0 & \frac{N_i}{r} & 0 \\ 0 & 0 & -\frac{dN_i}{ds} \\ 0 & 0 & -\frac{\sin\phi}{r} \\ -\sin\phi \frac{dN_i}{ds} & \cos\phi \frac{dN_i}{ds} & -N_i \end{bmatrix} \quad (i=1, 2)$$

$$\delta_i^e = [\bar{u}_i, \bar{w}_i, \beta_i] \quad (i = 1, 2)$$

并有

$$\frac{dN_1}{ds} = -\frac{1}{L}, \frac{dN_2}{ds} = \frac{1}{L}$$

至此可按标准步骤形成单元刚度矩阵和节点载荷向量,与薄壳截锥单元类似。

$$K^e = \int_0^1 2\pi r LB^T DB ds \quad P_i^e = 2\pi L T^T \int_0^1 N_i^T p r d\xi \quad (i = 1, 2)$$

### 2.3 曲边单元

曲边单元有 3 个节点,其形状、位移和坐标如图

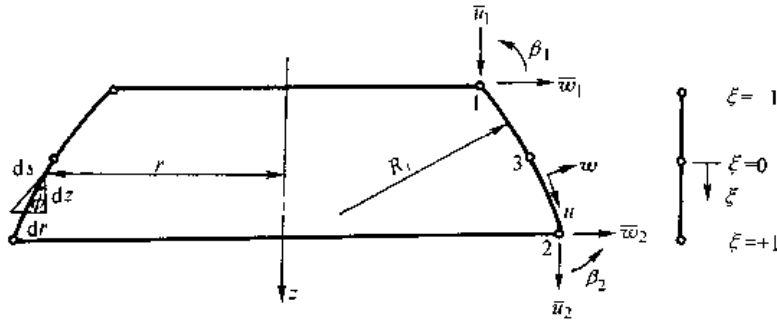


图 40.5-3 曲边单元

节点的位移向量为

$$\delta_i^e = [\bar{u}_i, \bar{w}_i, \beta_i]^T \quad (i = 1, 2, 3)$$

令

$$J = \frac{ds}{d\xi} = \sqrt{\left(\frac{dr}{d\xi}\right)^2 + \left(\frac{dz}{d\xi}\right)^2} = \sqrt{\left(\sum_{i=1}^3 \frac{dN_i}{d\xi} r_i\right)^2 + \left(\sum_{i=1}^3 \frac{dN_i}{d\xi} z_i\right)^2}$$

可得

$$\cos\phi = \frac{dz}{ds} = \frac{1}{J} \frac{dz}{d\xi} = \frac{1}{J} \sum_{i=1}^3 \frac{dN_i}{d\xi} z_i$$

$$\sin\phi = \frac{dr}{ds} = \frac{1}{J} \frac{dr}{d\xi} = \frac{1}{J} \sum_{i=1}^3 \frac{dN_i}{d\xi} r_i$$

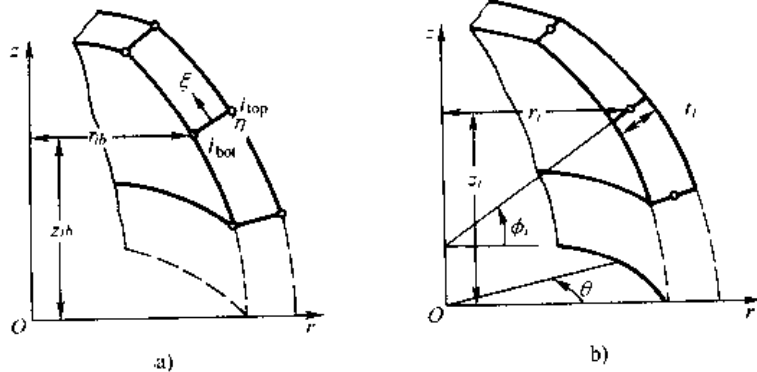


图 40.5-4 轴对称超参数单元

40.5-3 所示。壳体中面的坐标  $r, z$  和整体坐标系中的位移函数表示如下:

$$r = \sum_{i=1}^3 N_i r_i, z = \sum_{i=1}^3 N_i z_i$$

$$\bar{u} = \sum_{i=1}^3 N_i \bar{u}_i, \bar{w} = \sum_{i=1}^3 N_i \bar{w}_i,$$

$$\beta = \sum_{i=1}^3 N_i \beta_i$$

其中  $N_1 = (1-\xi)(1-2\xi), N_2 = \xi(2\xi-1),$   
 $N_3 = 4\xi(1-\xi)$

式中  $\xi$  是自然坐标,  $0 \leq \xi \leq 1$

进一步可以利用标准步骤计算单元刚度矩阵和载荷向量,各有关公式和截锥单元形式上相同,只是计算曲边单元时  $i=1, 2, 3$ ,同时各式中的  $L$  应代之以  $J$ ,并置于积分号内。

### 3 轴对称超参数壳体单元

#### 3.1 几何形状的规定

轴对称超参数壳体单元是从轴对称实体单元蜕化而来的,沿厚度方向位移为线性变化,单元如图 40.5-4 所示。

常用的线性、二次和三次的轴对称壳体单元的形状函数分别是

线性单元  $N_i = \frac{1}{2}(1 + \xi_i \xi)$   $\xi_i = \pm 1$

二次单元 端节点  $N_i = \frac{1}{2}(1 + \xi_i \xi)\xi_i \xi$

边中节点  $N_i = (1 - \xi^2)$

三次单元 端节点  $N_i = \frac{1}{16}(1 + \xi_i \xi)(-1 + 9\xi^2)$

边内节点  $N_i = \frac{1}{16}(1 + \xi_i \xi)(-1 + 9\xi^2)$

$\xi_i = \pm \frac{1}{3}$

### 3.2 位移函数

轴对称壳体内任一点的位移可由中面节点沿整体坐标  $r, z$  的位移分量  $u, v$  即法线的转动  $\alpha$  表示。在壳体单元内, 中面节点  $i$  的位移分量是  $u_i, v_i, \alpha_i$ 。单元内任一点的位移函数的表达式为

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = \begin{bmatrix} [N_1] & [N_2] & \dots & [N_n] \end{bmatrix} \begin{Bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{Bmatrix}$$

其中

$$\delta = [u_i \quad v_i \quad \alpha_i]^T$$

$$[N_i] = \begin{bmatrix} N_i & 0 & -N_i \eta \frac{t_i}{2} \sin \phi_i \\ 0 & N_i & N_i \eta \frac{t_i}{2} \cos \phi_i \end{bmatrix}$$

$(i = 1, 2, \dots, n)$

式中  $r, z$  —— 整体坐标;

$z$  —— 壳体的对称轴;

$\xi, \eta$  —— 自然坐标,  $-1 \leq \xi, \eta \leq 1, \eta$  沿中面法线方向;

$V_i$  —— 连接壳体单元下表面节点  $i_{\text{bot}}$  到上表面节点  $i_{\text{top}}$  的向量, 即

$$V_i = \begin{Bmatrix} V_{ir} \\ V_{iz} \end{Bmatrix} = \begin{Bmatrix} r_i \\ z_i \end{Bmatrix}_{\text{top}} - \begin{Bmatrix} r_i \\ z_i \end{Bmatrix}_{\text{bot}} = \begin{Bmatrix} \Delta r_i \\ \Delta z_i \end{Bmatrix}$$

$$t_i = \sqrt{(\Delta r_i)^2 + (\Delta z_i)^2}, \begin{Bmatrix} \cos \phi_i \\ \sin \phi_i \end{Bmatrix} = \begin{Bmatrix} \Delta r_i / t_i \\ \Delta z_i / t_i \end{Bmatrix}$$

### 3.3 应变和应力的确定

按壳体理论的基本假设, 在  $\eta$  为常数曲面的垂直方向应力应等于零, 因此应在以此方向为  $r'$  轴的局部坐标系  $r', z'$  内确定壳体的应力和应变。

在局部坐标系  $r', z'$  中应变分量的表达式是

$$\epsilon' = \begin{Bmatrix} \epsilon_{r'} \\ \epsilon_{\theta} \\ \gamma_{r'z'} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial u'}{\partial z'} \\ u' \\ r' \\ \frac{\partial u'}{\partial r'} + \frac{\partial v'}{\partial z'} \end{Bmatrix}$$

在计算  $\epsilon'$  中各个位移偏导数时, 首先将整体坐标系  $r, z$  和自然坐标  $\xi, \eta$  进行转换

$$\begin{bmatrix} \frac{\partial u}{\partial r} & \frac{\partial v}{\partial r} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} \end{bmatrix} = J^{-1} \begin{bmatrix} \frac{\partial u}{\partial \eta} & \frac{\partial v}{\partial \eta} \\ \frac{\partial u}{\partial \xi} & \frac{\partial v}{\partial \xi} \end{bmatrix}$$

其中  $J$  是 jacobian 矩阵

$$J = \begin{bmatrix} \frac{\partial r}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial r}{\partial \xi} & \frac{\partial z}{\partial \xi} \end{bmatrix}$$

其次将整体坐标系的各个位移偏导数转换为局部坐标系  $r', z'$  的位移偏导数

$$\begin{bmatrix} \frac{\partial u'}{\partial r'} & \frac{\partial v'}{\partial r'} \\ \frac{\partial u'}{\partial z'} & \frac{\partial v'}{\partial z'} \end{bmatrix} = T^T \begin{bmatrix} \frac{\partial u}{\partial r} & \frac{\partial v}{\partial r} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} \end{bmatrix} T$$

其中

$$T = \frac{1}{J(\eta)} \begin{bmatrix} \frac{\partial z}{\partial \xi} & \frac{\partial r}{\partial \xi} \\ -\frac{\partial r}{\partial \xi} & \frac{\partial z}{\partial \xi} \end{bmatrix}$$

式中  $J(\eta) = \sqrt{\left(\frac{\partial r}{\partial \xi}\right)^2 + \left(\frac{\partial z}{\partial \xi}\right)^2}$

最后可将  $\epsilon'$  表示成

$$\epsilon' = [B'_1 \quad B'_2 \quad \dots \quad B'_n] \begin{Bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{Bmatrix}$$

局部坐标系内的应力利用弹性关系可表示成

$$\sigma' = \begin{Bmatrix} \sigma_{r'} \\ \sigma_{\theta} \\ \tau_{r'z'} \end{Bmatrix} = D \epsilon'$$

其中

$$D = \frac{E}{1 - \mu^2} \begin{bmatrix} 1 & \mu & 0 \\ \mu & 1 & 0 \\ 0 & 0 & \frac{1 - \mu}{2k} \end{bmatrix}$$

式中  $E, \mu$  —— 材料常数;

$k$  —— 考虑切应力沿厚度不均匀分布引入的系数,  $k = 1, 2$ 。

### 3.4 刚度矩阵和节点载荷的计算

单元刚度矩阵可直接在自然坐标中用下式计算

$$\bar{K} = 2\pi \int_{-1}^1 \int_{-1}^1 B^T D B' |J| r d\eta d\xi$$

### 4 不同类型的单元的联结

在工程实际问题中,常常遇到三维连续体和薄壁板壳组成的结构。以下将给出就三维实体单元和壳体单元的联结的解决办法。

#### 4.1 多点约束方程

假设实体单元在交界面上有点 1、2、3,节点位移参数是  $u_1, v_1, u_2, v_2, u_3, v_3$ ;壳体单元的节点 2', 一般情况下可能有 3 个位移参数  $u_2', v_2', \beta_2'$ 。为了保证交界面上的协调性,得到如下各个位移协调条件的表达式

$$C = \begin{cases} \begin{cases} u_2 - u_2' \\ v_2 - v_2' \\ -\sin\phi(u_1 - u_2) + \cos\phi(v_1 - v_2) - \frac{t}{2}\beta_2' \end{cases} \\ \begin{cases} -\sin\phi(u_3 - u_2) + \cos\phi(v_3 - v_2) + \frac{t}{2}\beta_2' \end{cases} \end{cases}$$

在实际计算中,引入上列约束方程时有两种方案

$$\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin^2\phi & -\sin\phi\cos\phi & -\frac{t}{2}\sin\phi \\ \sin\phi & 0 & -\sin\phi\cos\phi & \cos^2\phi & \frac{t}{2}\cos\phi \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \cos\phi & \sin^2\phi & -\sin\phi\cos\phi & \frac{t}{2}\sin\phi \\ 0 & \sin\phi & -\sin\phi\cos\phi & \cos^2\phi & -\frac{t}{2}\cos\phi \end{bmatrix} \begin{Bmatrix} u_1' \\ v_1' \\ u_2' \\ v_2' \\ \beta_2' \end{Bmatrix}$$

直接引入法是将上列转换关系式引入实体部分的刚度矩阵和载荷向量,经转换后的实体部分的刚度矩阵和载荷向量可按通常的步骤,与壳体部分的矩阵或向量集成系统的刚度矩阵和载荷向量,从而得到系统的求解方程组。

#### 4.2 过渡单元

用于轴对称应力分析和三维应力分析的过渡单元实际上是轴对称等参实体单元和蜕化壳体单元(超参单元)的结合。利用轴对称等参实体单元和轴对称蜕化壳体单元的位移函数可得过渡单元的坐标变换和位移函数如下:

可供选择:

#### 4.1.1 罚函数法

通过罚数  $\alpha$  将约束方程引入系统的能量泛函

$$\pi^* = \pi - \frac{1}{2} \alpha C^T C$$

其中  $\pi$  是未考虑约束条件时系统的能量泛函,它是由实体单元和壳体单元两个区域能量泛函叠加而得到的。

由  $\delta\pi^* = 0$ ,可以得到

$$(K_1 + \alpha K_2)\delta = P$$

式中  $\delta, P$  ——系统的节点位移向量和载荷向量;

$K_1$  ——未引入约束方程时的系统刚度矩阵;

$K_2$  ——由于引入约束方程而增加的刚度矩阵。

#### 4.1.2 直接引入法

由于交界面上 9 个位移参数之间存在约束方程,选择  $u_1', u_3', u_2', v_2', \beta_2'$  作为独立位移参数,实体单元的 6 个位移参数  $u_1, v_1, u_2, v_2, u_3, v_3$  和它们之间存在以下转换关系:

$$\begin{Bmatrix} r \\ z \end{Bmatrix} = \sum_{i=1}^m N_i(\xi, \eta) \begin{Bmatrix} r_i \\ z_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi) \begin{Bmatrix} r_i \\ z_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi) \eta \frac{t_i}{2} \begin{Bmatrix} \cos\phi_i \\ \sin\phi_i \end{Bmatrix} a_i$$

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = \sum_{i=1}^m N_i(\xi, \eta) \begin{Bmatrix} u_i \\ v_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi) \begin{Bmatrix} u_i \\ v_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi) \eta \frac{t_i}{2} \begin{Bmatrix} \cos\phi_i \\ \sin\phi_i \end{Bmatrix} a_i$$

式中  $m$  ——过渡单元中和实体单元相关的节点数;  
 $n-m$  ——和壳体单元相关的节点数;  
 $n$  ——过渡单元的节点总数。

# 第 6 章 一般壳体问题的有限单元法

## 1 平板壳体单元

应用平板壳体单元进行有限单元法求解就是用折板代替壳体的计算方法。如图 40.6-1 所示,通常用三

角形或矩形薄板单元的组合去代替薄壳,其中因三角形单元可以适应壳体的复杂外形而有较广的应用,实用价值更大。

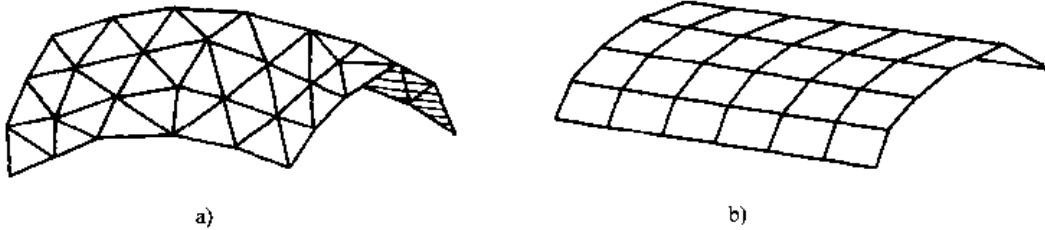


图 40.6-1 用折板代替薄壳

a) 由三角形单元组成的任意薄壳 b) 由矩形单元组成的棱柱面薄壳

### 1.1 局部坐标系内的单元刚度矩阵

平板壳体单元可以看成是平面应力单元和平板弯曲单元的组合,其单元刚度矩阵可以由这两种单元的刚度矩阵组合而成。

对于平面应力单元,其位移函数、节点位移参数和单元刚度矩阵为

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = \sum_{i=1}^n N_i^m \delta_i^m, \delta_i^m = \begin{Bmatrix} u_i \\ v_i \end{Bmatrix}, N_i^m = \begin{bmatrix} N_i & 0 \\ 0 & N_i \end{bmatrix}$$

$$K_{ij}^m = \iint (B_i^m)^T D^m B_j^m dx dy$$

式中  $N_i^m, B_i^m, D^m$  分别是平面应力单元的形函数矩阵、几何矩阵和弹性矩阵,上标  $m$  表示属于薄膜应力状态。

对于平板弯曲单元,其位移函数、节点位移参数和单元刚度矩阵为

$$w = \sum_{i=1}^n N_i^b \delta_i^b, \delta_i^b = \begin{Bmatrix} w_i \\ \theta_{xi} \\ \theta_{yi} \end{Bmatrix}, N_i^b = \begin{bmatrix} N_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & N_i \end{bmatrix}$$

$$K_{ij}^b = \iint (B_i^b)^T D^b B_j^b dx dy$$

式中上标  $b$  表示属于平板弯曲状态。

组合上述两种状态得到平板壳体单元的各个矩阵表达式。因将局部坐标系的刚度矩阵转换到整体坐标系进行集成的需要,将  $\theta_z$  补充到节点位移参数中,得到平板壳体单元的节点位移向量

$$\delta_i = [u_i, v_i, w_i, \theta_{xi}, \theta_{yi}, \theta_{zi}]^T$$

平板壳体单元的刚度矩阵可表示为

$$\bar{K}_{ij}^c = \begin{bmatrix} \bar{K}_{ij}^m & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ 0 & 0 & & & 0 \\ 0 & 0 & & \bar{K}_{ij}^b & 0 \\ 0 & 0 & & & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

### 1.2 单元刚度矩阵从局部坐标系到整体坐标系的转换

局部坐标系内的节点位移向量是

$$\delta_i = [u_i, v_i, w_i, \theta_{xi}, \theta_{yi}, \theta_{zi}]^T$$

整体坐标系内的节点位移向量是

$$\delta_i' = [u_i', v_i', w_i', \theta_{xi}', \theta_{yi}', \theta_{zi}']^T$$

节点位移向量在两个坐标系之间的转换关系式是

$$\delta_i' = T \delta_i$$

$$T = \begin{bmatrix} L & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & L \end{bmatrix},$$

$$L = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix},$$

$$\lambda = \begin{bmatrix} \lambda_{x'x} & \lambda_{x'y} & \lambda_{x'z} \\ \lambda_{y'x} & \lambda_{y'y} & \lambda_{y'z} \\ \lambda_{z'x} & \lambda_{z'y} & \lambda_{z'z} \end{bmatrix}$$

式中  $\lambda_{x'x} = \cos(x', x)$  等是  $x, y, z$  轴在  $x', y', z'$  系的各个方向余弦。

单元刚度矩阵和载荷向量的转换关系是

$$K^{i'c} = T \bar{K}^c T^T, P^{i'c} = T P^c$$

## 2 超参数壳体单元

### 2.1 几何形状的规定

典型的厚壳单元,由上、下两个曲面及周边以壳体厚度方向的直线为母线的曲面所围成。在厚壳单元的中面上规定曲线坐标  $\xi, \eta$ , 厚度方向的直线坐标  $\zeta$ , 且  $-1 \leq \xi, \eta, \zeta \leq 1$ , 如图 40.6-2 所示。

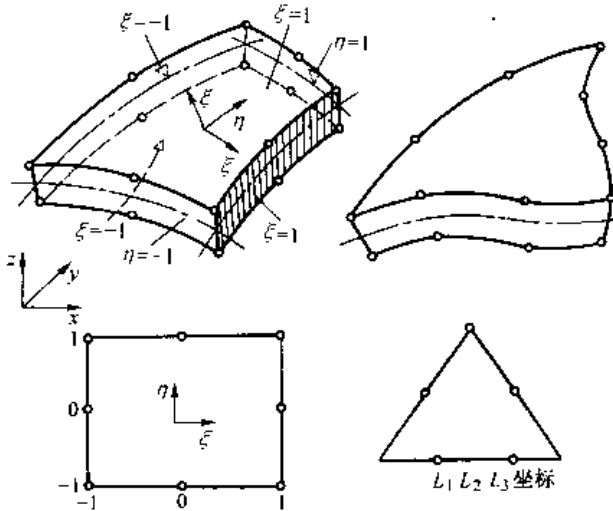


图 40.6-2 不同形状的厚壳单元及其局部坐标

### 2.2 位移函数的表示

在每个节点  $i$  有 3 个线位移和 2 个角位移。为了定义角位移,过节点  $i$  作 3 个正交向量  $v_1, v_2, v_3$ 。  $v_3$  在中面的法线方向,  $v_1$  和  $v_2$  与  $v_3$  垂直并相互正交,而且相切于中面。  $v_3$  绕  $v_1$  和  $v_2$  的旋转分别为  $\beta_1$  和  $\beta_2$ , 则单元的节点位移向量

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = \begin{bmatrix} [N_1] & [N_2] & \dots & [N_n] \end{bmatrix} \begin{Bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{Bmatrix}$$

其中

$$\delta_i = [u_i, v_i, w_i, \beta_1, \beta_2]^T$$

$$N_i = \begin{bmatrix} N_i & 0 & 0 & N_i \xi \frac{l_i}{2} l_{1i} & -N_i \xi \frac{l_i}{2} l_{2i} \\ 0 & N_i & 0 & N_i \xi \frac{l_i}{2} m_{1i} & -N_i \xi \frac{l_i}{2} m_{2i} \\ 0 & 0 & N_i & N_i \xi \frac{l_i}{2} n_{1i} & -N_i \xi \frac{l_i}{2} n_{2i} \end{bmatrix}$$

$(i = 1, 2, \dots, n)$

式中

$$l_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2 + (\Delta z_i)^2}$$

$l_i$  即为  $i$  节点的厚度;  $l_{1i}, m_{1i}, n_{1i}$  和  $l_{2i}, m_{2i}, n_{2i}$  分别是  $v_1$  和  $v_2$  的方向余弦。

### 2.3 应变和应力的确定

在局部坐标系  $x', y', z'$  中应变分量的表达式是

$$\epsilon' = \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial u'}{\partial x'} \\ \frac{\partial v'}{\partial y'} \\ \frac{\partial u'}{\partial y'} + \frac{\partial v'}{\partial x'} \\ \frac{\partial v'}{\partial z'} + \frac{\partial w'}{\partial y'} \\ \frac{\partial u'}{\partial z'} + \frac{\partial w'}{\partial x'} \end{Bmatrix}$$

在计算  $\epsilon'$  中各个位移偏导数时,首先将整体坐标系  $x, y, z$  和自然坐标  $\xi, \eta, \zeta$  进行转换

$$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} & \frac{\partial w}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} & \frac{\partial w}{\partial y} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} & \frac{\partial w}{\partial z} \end{bmatrix} = J^{-1} \begin{bmatrix} \frac{\partial u}{\partial \xi} & \frac{\partial v}{\partial \xi} & \frac{\partial w}{\partial \xi} \\ \frac{\partial u}{\partial \eta} & \frac{\partial v}{\partial \eta} & \frac{\partial w}{\partial \eta} \\ \frac{\partial u}{\partial \zeta} & \frac{\partial v}{\partial \zeta} & \frac{\partial w}{\partial \zeta} \end{bmatrix}$$

其中  $J$  是 jacobian 矩阵

$$J = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{bmatrix}$$

其次将整体坐标系的各个位移偏导数转换为局部坐标系  $x', z'$  的位移偏导数

$$\begin{bmatrix} \frac{\partial u'}{\partial x'} & \frac{\partial v'}{\partial x'} & \frac{\partial w'}{\partial x'} \\ \frac{\partial u'}{\partial y'} & \frac{\partial v'}{\partial y'} & \frac{\partial w'}{\partial y'} \\ \frac{\partial u'}{\partial z'} & \frac{\partial v'}{\partial z'} & \frac{\partial w'}{\partial z'} \end{bmatrix} = T^T \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} & \frac{\partial w}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} & \frac{\partial w}{\partial y} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} & \frac{\partial w}{\partial z} \end{bmatrix} T$$

其中

$$T = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix}$$

式中  $l_i, m_i, n_i$  是  $v_i$  的方向余弦 ( $i=1, 2, 3$ )。

最后可将  $\epsilon'$  表示成

$$\epsilon^t = [B'_1 \quad B'_2 \quad \dots \quad B'_n] \begin{Bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{Bmatrix}$$

局部坐标系内的应力利用弹性关系可表示成

$$\sigma^t = [\sigma_x \quad \sigma_y \quad \tau_{xy} \quad \tau_{yz} \quad \tau_{xz}]^T = D \epsilon^t$$

$$D = \frac{E}{1-\mu^2} \begin{bmatrix} 1 & \mu & 0 & 0 & 0 \\ \mu & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\mu}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1-\mu}{2k} & 0 \\ 0 & 0 & 0 & 0 & \frac{1-\mu}{2k} \end{bmatrix}$$

式中  $E, \mu$  — 材料常数;  
 $k$  — 考虑切应力沿厚度不均匀分布引入的系数,  $k=1, 2$ 。

### 2.4 刚度矩阵和节点载荷的计算

单元刚度矩阵可直接在自然坐标中用下式计算

$$\bar{K}^e = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 B'^t D B' |J| d\xi d\eta d\zeta$$

### 3 相对自由度壳体单元

相对自由度壳体单元就是对绝对坐标  $x, y, z$ , 和绝对位移  $u, v, w$ , 进行线性组合以定义新的相对坐标  $\bar{x}, \bar{y}, \bar{z}$ , 和相对位移  $\bar{u}, \bar{v}, \bar{w}$ 。以 16 节点三维等参元为例, 如图 40.6-3 所示, 它们之间的关系如下:

$$\left. \begin{aligned} \bar{x}_i &= \frac{1}{2}(x_i - x_{i-1}) \\ \bar{x}_{i+1} &= \frac{1}{2}(x_i + x_{i+1}) \end{aligned} \right\} (i=1, 2, 3, 4, 9, 10, 11, 12)$$

$$\left. \begin{aligned} \bar{u}_i &= \frac{1}{2}(u_i - u_{i+1}) \\ \bar{u}_{i+1} &= \frac{1}{2}(u_i + u_{i+1}) \end{aligned} \right\} (i=1, 2, 3, 4, 9, 10, 11, 12)$$

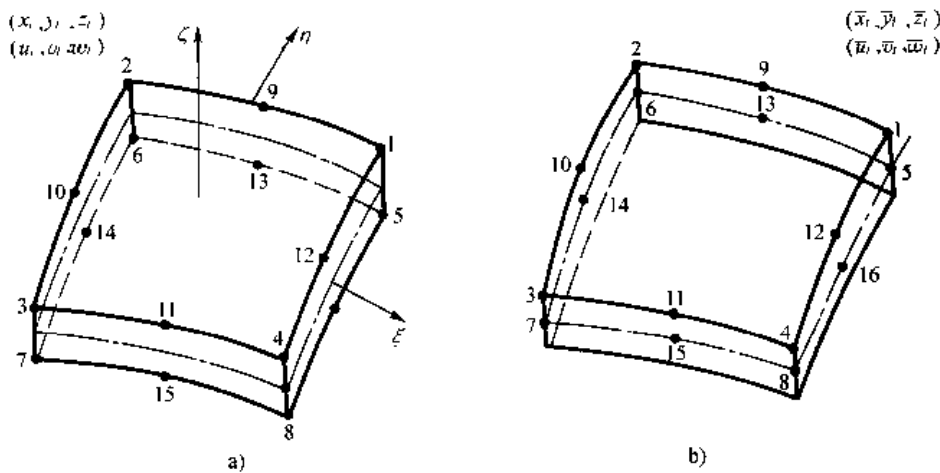


图 40.6-3 二维等参元和相对自由度壳元

其余坐标  $\bar{y}, \bar{z}$ , 及位移  $\bar{v}, \bar{w}$ , 具有类似表达式。用相对自由度壳体单元的节点坐标和节点位移表示原等参元的节点坐标和节点位移, 则有

$$\left. \begin{aligned} x_i &= \bar{x}_i + \bar{x}_{i+1} \\ x_{i+1} &= \bar{x}_{i-1} - \bar{x}_i \\ u_i &= \bar{u}_i - \bar{u}_{i+1} \\ u_{i+1} &= \bar{u}_{i+1} - \bar{u}_i \end{aligned} \right\} (i=1, 2, 3, 4, 9, 10, 11, 12)$$

对于坐标  $y, z$ , 及位移  $v, w$ , 具有类似表达式。同样可以得到坐标变换和位移函数

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \sum_{i=1}^{16} N_i \begin{Bmatrix} \bar{x}_i \\ \bar{y}_i \\ \bar{z}_i \end{Bmatrix}$$

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = \sum_{i=1}^{16} \bar{N}_i \begin{Bmatrix} \bar{u}_i \\ \bar{v}_i \\ \bar{w}_i \end{Bmatrix}$$

其中

$$\left. \begin{aligned} \bar{N}_i &= N_i - N_{i-4} \\ \bar{N}_{i+4} &= N_i + N_{i+4} \end{aligned} \right\} (i = 1, 2, 3, 4, 9, 10, 11, 12)$$

### 4 不同类型单元的联结

#### 4.1 多点约束方程

在三维等参实体单元和超参壳体单元相联结时，如果在交界面上，和壳体单元的节点  $i$  相对应，在实体单元上有三个节点  $i_t, i_m, i_b$ ，它们分别布置在顶面、中面和底面上，在局部坐标系  $v_1, v_2, v_3$  中，建立两个单

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \sum_{i=1}^m N_i(\xi, \eta, \zeta) \begin{Bmatrix} x_i \\ y_i \\ z_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi, \eta) \begin{Bmatrix} x_i \\ y_i \\ z_i \end{Bmatrix}_{\text{中面}} + \frac{\zeta}{2} \begin{Bmatrix} \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \eta} - \frac{\partial y}{\partial \eta} \frac{\partial x}{\partial \xi} \\ \frac{\partial z}{\partial \xi} \frac{\partial x}{\partial \eta} - \frac{\partial z}{\partial \eta} \frac{\partial x}{\partial \xi} \\ \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta} \frac{\partial y}{\partial \xi} \end{Bmatrix}$$

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = \sum_{i=1}^m N_i(\xi, \eta, \zeta) \begin{Bmatrix} u_i \\ v_i \\ w_i \end{Bmatrix} + \sum_{i=m+1}^n N_i(\xi, \eta) \begin{Bmatrix} u_i \\ v_i \\ w_i \end{Bmatrix}_{\text{中面}} + \frac{\zeta}{2} \begin{bmatrix} l_{1i} & -l_{2i} \\ m_{1i} & -m_{2i} \\ n_{1i} & -n_{2i} \end{bmatrix} \begin{Bmatrix} \alpha_i \\ \beta_i \end{Bmatrix}$$

式中  $m$ ——过渡单元中和实体单元相关的节点数；  
 $n-m$ ——和壳体单元相关的节点数；

元节点参数之间的约束方程如下

$$C' = \begin{Bmatrix} u'_i - u'_{i_m} \\ v'_i - v'_{i_m} \\ w'_i - w'_{i_m} \\ u'_{i_m} - \frac{u'_{i_t} + u'_{i_b}}{2} \\ v'_{i_m} - \frac{v'_{i_t} + v'_{i_b}}{2} \\ \beta'_{i_m} - \frac{v'_{i_b} - v'_{i_t}}{l_i} \\ \alpha'_{i_m} - \frac{u'_{i_b} - u'_{i_t}}{l_i} \end{Bmatrix} = 0$$

#### 4.2 过渡单元

构造三维过渡单元的原则和方法和轴对称的情况相同，由等参实体单元和超参壳体单元结合而成的过渡单元的位移函数表示如下

$n$ ——过渡单元的节点总数。



## 第 7 章 热传导问题的有限单元法

在一般三维问题中,瞬态温度场的场变量  $\phi(x, y, z, t)$  在直角坐标系中应满足的微分方程是

$$\rho c \frac{\partial \phi}{\partial t} - \frac{\partial}{\partial x} \left( k_x \frac{\partial \phi}{\partial x} \right) - \frac{\partial}{\partial y} \left( k_y \frac{\partial \phi}{\partial y} \right) - \frac{\partial}{\partial z} \left( k_z \frac{\partial \phi}{\partial z} \right) - \rho Q = 0 \quad (\text{在 } \Omega \text{ 内})$$

边界条件是

$$\phi = \bar{\phi} \quad (\text{在 } \Gamma_1 \text{ 边界上})$$

$$k_x \frac{\partial \phi}{\partial x} n_x + k_y \frac{\partial \phi}{\partial y} n_y + k_z \frac{\partial \phi}{\partial z} n_z = q \quad (\text{在 } \Gamma_2 \text{ 边界上})$$

$$k_x \frac{\partial \phi}{\partial x} n_x + k_y \frac{\partial \phi}{\partial y} n_y + k_z \frac{\partial \phi}{\partial z} n_z = h(\phi_a - \phi) \quad (\text{在 } \Gamma_3 \text{ 边界上})$$

式中  $\rho$  材料密度 ( $\text{kg}/\text{m}^3$ );

$c$ ——材料比热 [ $\text{J}/(\text{kg} \cdot \text{K})$ ];

$t$ ——时间 ( $\text{s}$ );

$k_x, k_y, k_z$ ——材料沿  $x, y, z$  方向的热传导系数 [ $\text{W}/(\text{m} \cdot \text{K})$ ];

$Q$ ——物体内部的热源密度 ( $\text{W}/\text{kg}$ );

$$Q = Q(x, y, z, t)$$

$n_x, n_y, n_z$ ——边界外法线的方向余弦;

$\bar{\phi}$ —— $\Gamma_1$  边界上的给定温度;

$$\bar{\phi} = \bar{\phi}(\Gamma, t)$$

$q$ —— $\Gamma_2$  边界上的给定热流量 ( $\text{W}/\text{m}^2$ );

$$q = q(\Gamma, t)$$

$h$ ——放热系数 ( $\text{W}/\text{m}^2 \cdot \text{K}$ );

$\phi_a$  在自然对流条件下,是外界环境温度;在强迫对流条件下,是边界层的绝热壁温度。

$$\phi_a = \phi_a(\Gamma, t)$$

边界应满足

$$\Gamma_1 + \Gamma_2 + \Gamma_3 = \Gamma$$

其中  $\Gamma$  是  $\Omega$  域的全部边界。

如果边界上的  $\bar{\phi}, q, \phi_a$  及内部的  $Q$  不随时间变化,则经过一定时间的热交换后,物体内的各点温度也将不再随时间而变化,即

$$\frac{\partial \phi}{\partial t} = 0$$

这时瞬态热传导方程就退化为稳态热传导方程,则三维稳定热传导方程

$$\frac{\partial}{\partial x} \left( k_x \frac{\partial \phi}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial \phi}{\partial y} \right) + \frac{\partial}{\partial z} \left( k_z \frac{\partial \phi}{\partial z} \right) + \rho Q = 0 \quad (\text{在 } \Omega \text{ 内})$$

相应的二维稳定热传导方程

$$\frac{\partial}{\partial x} \left( k_x \frac{\partial \phi}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial \phi}{\partial y} \right) + \rho Q = 0 \quad (\text{在 } \Omega \text{ 内})$$

求解稳态温度场的问题就是求满足稳态热传导方程及边界条件的场变量  $\phi$ ,  $\phi$  只是坐标的函数,与时间无关。

### 1 稳态热传导问题

以下以二维单元为例,说明用伽辽金法建立稳定热传导有限元的一般格式。

构造近似场函数  $\tilde{\phi}$ , 并设  $\tilde{\phi}$  已满足  $\Gamma_1$  边界上的强制边界条件。将近似函数代入二维问题的稳定热传导方程及  $\Gamma_2$  边界和  $\Gamma_3$  边界的边界条件。因  $\tilde{\phi}$  的近似性,产生余量,即

$$\begin{cases} R_a = \frac{\partial}{\partial x} \left( k_x \frac{\partial \tilde{\phi}}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial \tilde{\phi}}{\partial y} \right) + \rho Q \\ R_{\Gamma_2} = k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - q \\ R_{\Gamma_3} = k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - h(\tilde{\phi}_a - \tilde{\phi}) \end{cases}$$

用加权余量法建立有限元格式的基本思想是使余量的加权积分为零,即

$$\int_{\Omega} R_a w_1 d\Omega + \int_{\Gamma_2} R_{\Gamma_2} w_2 d\Gamma + \int_{\Gamma_3} R_{\Gamma_3} w_3 d\Gamma = 0$$

式中  $w_1, w_2, w_3$ ——权函数。

将各个余量式代入加权积分的和式,并进行分部积分可得

$$\begin{aligned} & - \int_{\Omega} \left[ \frac{\partial w_1}{\partial x} \left( k_x \frac{\partial \tilde{\phi}}{\partial x} \right) + \frac{\partial w_1}{\partial y} \left( k_y \frac{\partial \tilde{\phi}}{\partial y} \right) - \rho Q w_1 \right] d\Omega \\ & + \int_{\Gamma_2} w_2 \left( k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y \right) d\Gamma \\ & + \int_{\Gamma_3} \left( k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - q \right) w_3 d\Gamma \end{aligned}$$

$$+ \int_{\Gamma_3} \left( k_r \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - h(\phi_a - \tilde{\phi}) \right) w_3 d\Gamma = 0$$

将空间域  $\Omega$  离散为有限个单元体, 在典型单元内各点的温度  $\phi$  可以近似的用单元的节点温度  $\phi$  插值得到

$$\begin{aligned} \phi &= \sum_{i=1}^n N_i(x, y) \phi_i \quad N\phi \\ N &= [N_1 \quad N_2 \quad \dots \quad N_n] \end{aligned}$$

式中  $n$  是每个单元的节点个数;  $N(x, y)$  是插值函数, 它是  $C_0$  型插值函数, 具有下述性质

$$N_i(x, y) = \begin{cases} 0, & (\text{当 } j \neq i) \\ 1, & (\text{当 } j = i) \end{cases} \quad \text{且} \quad \sum N_i = 1$$

用伽辽金选择权函数

$$w_j = N_j \quad (j=1, 2, \dots, n)$$

其中是域全部离散得到的节点总数。在边界上不失一般性地选择

$$w_2 = w_3 = -w_1 = -N_1 \quad (j=1, 2, \dots, n)$$

将插值函数代入加权积分的和式, 可以得到

$$\begin{aligned} &\sum_j \int_{\Omega} \left[ \left( \frac{\partial N}{\partial x} \right)^T k_x \frac{\partial N}{\partial x} + \left( \frac{\partial N}{\partial y} \right)^T k_y \frac{\partial N}{\partial y} \right] \phi d\Omega \\ &+ \sum_j \int_{\Gamma_1} h N^T N \phi d\Gamma \\ &- \sum_j \int_{\Gamma_2} N^T h \phi_a d\Gamma - \sum_j \int_{\Gamma} N^T \rho Q d\Omega = 0 \end{aligned}$$

按照一般有限元格式可表示如下

$$K\phi = P$$

式中  $K$  — 热传导矩阵;

$\phi = [\phi_1, \phi_2, \dots, \phi_n]^T$  — 节点温度列阵;

$P$  — 温度载荷列阵。

热传导矩阵和温度载荷列阵的单元集成形式为

$$K_{ij} = \sum_e K_{ij}^e + \sum_e H_{ij}^e$$

$$P_i = \sum_e P_{q_i}^e + \sum_e P_{h_i}^e + \sum_e P_{\phi_a}^e$$

式中

$$K_{ij}^e = \int_{\Omega} \left( k_x \frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + k_y \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} \right) d\Omega$$

$$H_{ij}^e = \int_{\Gamma_1} h N_i N_j d\Gamma$$

$$P_{q_i}^e = \int_{\Omega} N_i \rho Q d\Omega$$

$$P_{h_i}^e = \int_{\Gamma_2} N_i h \phi_a d\Gamma$$

$$P_{\phi_a}^e = \int_{\Gamma_3} N_i \rho c Q d\Omega$$

## 2 瞬态热传导问题

瞬态温度场与稳态温度场的主要差别是瞬态温度场的场函数不仅是空间域  $\Omega$  的函数, 而且还是时间域  $t$  的函数。在瞬态热传导问题中, 节点温度  $\phi$  是时间  $t$  的函数, 则典型单元内的温度  $\phi$  可近似的表示为

$$\phi = \tilde{\phi} = \sum_{i=1}^n N_i(x, y) \phi_i(t)$$

插值函数  $N_i$  只是空间域的函数, 与稳态热传导问题相同。将近似函数代入二维问题的稳定热传导方程及  $\Gamma_2$  边界和  $\Gamma_3$  边界的边界条件。因  $\tilde{\phi}$  的近似性, 产生余量, 即

$$\begin{cases} R_{\Omega} = \frac{\partial}{\partial x} \left( k_x \frac{\partial \tilde{\phi}}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial \tilde{\phi}}{\partial y} \right) + \rho Q - \rho c \frac{\partial \tilde{\phi}}{\partial t} \\ R_{\Gamma_2} = k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - q \\ R_{\Gamma_3} = k_x \frac{\partial \tilde{\phi}}{\partial x} n_x + k_y \frac{\partial \tilde{\phi}}{\partial y} n_y - h(\phi_a - \tilde{\phi}) \end{cases}$$

令余量的加权积分为零, 即

$$\int_{\Omega} R_{\Omega} w_1 d\Omega - \int_{\Gamma_2} R_{\Gamma_2} w_2 d\Gamma + \int_{\Gamma_3} R_{\Gamma_3} w_3 d\Gamma = 0$$

按伽辽金法选择权函数

$$w_2 = w_3 = -w_1 = -N_1 \quad (j=1, 2, \dots, n)$$

再由余量积分的和式, 经分部积分, 得一般有限元格式的矩阵方程

$$C\dot{\phi} + K\phi = P$$

式中  $C$  — 热容矩阵;

$K$  — 热传导矩阵;

$\phi = [\phi_1, \phi_2, \dots, \phi_n]^T$  — 节点温度列阵;

$\dot{\phi}$  — 节点温度对时间的导数列阵,  $\dot{\phi} = d\phi/dt$ ;

$P$  — 温度载荷列阵。

热传导矩阵和温度载荷列阵的单元集成形式为

$$K_{ij} = \sum_e K_{ij}^e + \sum_e H_{ij}^e$$

$$C_{ij} = \sum_e C_{ij}^e$$

$$P_i = \sum_e P_{q_i}^e + \sum_e P_{h_i}^e + \sum_e P_{\phi_a}^e$$

$$K_{ij}^e = \int_{\Omega} \left( k_x \frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + k_y \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} \right) d\Omega$$

$$H_{ij}^e = \int_{\Gamma_1} h N_i N_j d\Gamma$$

$$C_{ij}^e = \int_{\Omega} \rho c N_i N_j d\Omega$$

$$P_{q_i}^e = \int_{\Omega} N_i \rho Q d\Omega$$

$$P_{H_i} = \int_{\Gamma_i^e} N_i h \phi_e d\Gamma$$

$$P_{Q_i} = \int_{\Omega^e} N_i \rho Q d\Omega$$

### 3 热应力的计算

当物体各部分温度发生变化时,物体由于热变形而产生线应变  $\alpha(\phi - \phi_0)$ ,切应变为零。这种由于热变形产生的应变可以看作是物体的初应变。计算热应力时只需算出热变形引起的初应变  $\epsilon_0$ ,求得相应的初应变引起的等效节点载荷  $P_{\epsilon_0}$ (简称温度载荷),然后按通常

求解应力的过程解得由于热变形引起的节点位移  $\phi$ ,然后由位移  $\phi$ 求得热应力  $\sigma$ 。

对于三维问题,温度变化引起的温度应变的表达式是

$$\epsilon_0 = \alpha(\phi - \phi_0)[1 \ 1 \ 1 \ 0 \ 0 \ 0]^T$$

式中  $\alpha$ ——材料的热膨胀系数( $1/^\circ\text{C}$ );

$\phi_0$ ——结构的初始温度场。

由温度应变引起的载荷项  $P_{\epsilon_0}$ 是

$$P_{\epsilon_0} = \sum_r \int_{\Omega^e} \mathbf{B}^T D \epsilon_0 d\Omega$$

## 第 8 章 动力学问题的有限单元法

弹性系统离散化以后,系统的运动方程为

$$M\ddot{\delta}(t) + C\dot{\delta}(t) + K\delta(t) = P(t)$$

式中  $\ddot{\delta}(t), \dot{\delta}(t)$ ——系统的节点加速度向量和节点速度向量;  
 $M, C, K, P(t)$ ——系统的质量矩阵、阻尼矩阵、刚度矩阵和节点载荷向量。  
 分别由各自的单元矩阵和向量集成

$$M = \sum_e M^e, C = \sum_e C^e,$$

$$K = \sum_e K^e, P = \sum_e P^e$$

$$M^e = \int_{V_e} \rho N^T N dV, C^e = \int_{V_e} c N^T N dV,$$

$$K^e = \int_{V_e} B^T D B dV$$

分别是单元的质量矩阵、刚度矩阵和阻尼矩阵,  $\rho$  是系统的密度,  $c$  是阻尼系数。

$$P^e = \int_{V_e} N^T f dV + \int_{S_e} N^T q dS$$

是单元载荷向量,  $f, q$  分别为单元的分布体积力和分布面力。

忽略阻尼的影响,则系统的运动方程简化为

$$M\ddot{\delta}(t) + K\delta(t) = P(t)$$

### 1 质量矩阵与阻尼矩阵

#### 1.1 协调质量矩阵和集中质量矩阵

1) 协调质量矩阵(一致质量矩阵) 从单元的动能导出,质量分布按照实际分布情况,同时位移插值函数和从位能导出刚度矩阵所采用的形式相同。其表达式为:

$$M^e = \int_{V_e} \rho N^T N dV$$

2) 集中质量矩阵(团聚质量矩阵) 假定单元的质量集中在节点上,单元质量矩阵为对角阵。  
 对于平面应力和应变单元,以三角形为例,其单元的协调质量矩阵为:

$$M^e = \frac{W}{3} \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix}$$

式中  $W = \rho t \Delta$ , 是单元的质量,  $t$  是单元的密度,  $\Delta$  是三角形单元的面积。

其单元的集中质量矩阵为:

$$M^e = \frac{W}{3} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

#### 1.2 振型阻尼矩阵

阻尼力正比于质点运动速度的单元阻尼矩阵表示为

$$C^e = \int_{V_e} c N^T N dV$$

上式的阻尼矩阵称为协调阻尼矩阵,通常均将介质阻尼简化为这种情况。这时的阻尼比例于单元质量矩阵。

阻尼力比例于应变速度的阻尼,例如由于材料内摩擦引起的结构阻尼,这时阻尼力可以表示成  $c D \dot{\epsilon}$ , 则单元阻尼矩阵表示为

$$C^e = \int_{V_e} c B^T D B dV$$

此单元阻尼矩阵比例于单元刚度矩阵。

在实际分析中,要精确地决定阻尼矩阵是相当困难的,通常允许将实际结构的阻尼矩阵简化为  $M$  和  $K$  的线性组合,即

$$C = \alpha M + \beta K$$

其中  $\alpha, \beta$  是不依赖于频率的常数。这种振型阻尼称为 Rayleigh 阻尼。

### 2 直接积分法

直接积分法是指在积分运动方程之前不进行方程形式的变换,而直接进行逐步数值积分。



其中

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$$

将以上变换代入到系统的运动方程,可得到新基向量空间内的运动方程

$$\mathbf{x}''(t) + \Phi^T C \Phi \mathbf{x}'(t) + \Omega^2 \mathbf{x}(t) = \Phi^T P(t)$$

初始条件也相应地转换成

$$\mathbf{x}_0 = \Phi^T M \dot{\delta}_0 \quad \mathbf{x}_1 = \Phi^T M \delta_0$$

### 3.2 求解单自由度系统振动方程

单自由度系统振动方程的求解可以应用直接积分方法和杜哈美积分。杜哈美积分又称为叠加积分,其基本思想是将任意激振力分解为一系列微冲量的连续作用,分别求出系统对每个微冲量的响应,然后根据线性叠加原理,将它们叠加起来,得到系统对任意激振的响应。杜哈美积分的结果是

$$x_1(t) = \frac{1}{\omega_1} \int_0^t r_1(\tau) e^{-\xi_1 \bar{\omega}_1 (t-\tau)} \sin \bar{\omega}_1 (t-\tau) d\tau + e^{-\xi_1 \bar{\omega}_1 t} (a_1 \sin \bar{\omega}_1 t + b_1 \cos \bar{\omega}_1 t)$$

式中  $\bar{\omega}_1 = \omega_1 \sqrt{1 - \xi_1^2}$ ;

$a_1, b_1$  —— 由起始条件决定的;

$\xi_i (i=1, 2, \dots, n)$  —— 第  $i$  阶振型阻尼比。

### 3.3 振型叠加得到系统的响应

在得到每个振型的响应之后,将它们叠加起来就得到系统的响应,即

$$\delta(t) = \sum_{i=1}^n \phi_i r_i(t)$$

## 4 大型特征值问题的解法

应用较为广泛的效率较高的特征值问题的解法主要是矩阵反迭代法和子空间迭代法。前者算法简单,比较适合于只要求得到系统的很少数目特征解的情况。后者实质是将前者推广与同时利用若干个向量进行迭代的情况,可以用于要求得到系统稍多一些特征解的情况。

### 4.1 反迭代法

利用反迭代法求解广义特征值问题是依次逐个求解特征解  $(\omega_1^2, \phi_1), (\omega_2^2, \phi_2), \dots$ 。

### 4.2 子空间迭代法

子空间迭代法是求解大型矩阵特征值问题的最有效方法之一,适合于求解部分特征解,广泛应用于结构动力学的有限元分析中。

子空间迭代法是假设  $r$  个起始向量同时进行迭代,以求得矩阵的前  $r$  个特征值和特征向量。

### 4.3 Ritz 向量直接叠加法

Ritz 向量直接叠加法的基本点是,根据载荷空间分布模式按一定规律生成一组 Ritz 向量,在将系统的运动方程转换到这组 Ritz 向量空间以后,只要求解一次减缩了的标准特征值问题,再经过坐标变换,就可以得到原系统的运动方程的部分特征解。

## 5 减缩系统自由度的方法

缩减系统自由度数目是广泛使用的求解动力学响应问题的方法。主要有主从自由度法和模态综合法。

### 5.1 主从自由度法

主从自由度法中,将根据刚度矩阵要求划分的网格总自由度,即位移向量  $\delta$ ,分为  $\delta_m$  和  $\delta_s$  两部分,并假定  $\delta_s$  按照一种确定的方法依赖于  $\delta_m$ 。因此  $\delta_m$  成为主自由度,而  $\delta_s$  称为从自由度。所以有:

$$\delta = \begin{bmatrix} I \\ T \end{bmatrix} \delta_m = T' \delta_m$$

其中  $\delta_s = T \delta_m$ ,  $T'$  规定了  $\delta_s$  和  $\delta_m$  之间的依赖关系。

采用上式建立的  $\delta_s$  和  $\delta_m$  之间的关系,实质上是假定对应于  $\delta_s$  自由度上的惯性力项已按静力等效原则转移到  $\delta_m$  自由度上。这只是当对应于这些自由度质量较小,而刚度较大,以及频率较低时才认为合理。随着频率的升高,误差也将增大,所以采用主从自由度法时,通常不宜分析高阶的频率和振型。

### 5.2 模态综合法

模态综合法分析实际结构的主要步骤:

(1) 将总体结构分割为若干子结构

依照结构的自然特点和分析的方便,将结构分成若干子结构,各个子结构通过交界面上的节点相互联结。

(2) 子结构的模态分析

仍以节点位移为基向量(简称物理坐标)建立子结构的运动方程

$$M^s \delta^s(t) + C^s \dot{\delta}^s(t) + K^s \delta^s(t) = P^s(t) + R^s(t)$$

式中 上标  $s$  —— 该矩阵或向量是属于子结构  $s (s=1, 2, \dots, r)$  的;

$r$  —— 子结构数;

$P^s(t)$  —— 外载荷向量;

$R^s(t)$  —— 交界面上的力向量。

而后对每个子结构按照求解运动方程的一般步骤求解。

(3) 综合各子结构的运动方程得到整个结构系统的运动方程并求解。

(4) 由模态坐标返回到各子结构的物理坐标。

## 第 9 章 材料非线性问题的有限单元法

### 1 材料非线性本构关系

在初始弹性范围内, 应力与应变之间存在着——对应的关系, 即广义胡克定律。进入塑性状态后, 一般来说, 不再存在着应力与应变之间的——对应关系, 只能建立应力增量与应变增量之间的关系。这种用增量形式表示的材料本构关系, 称为增量理论或流动理论。

在小应变的情况下, 应变增量可以分为弹性和塑性两部分, 即

$$d\epsilon_{ij} = d\epsilon_{ij}^e + d\epsilon_{ij}^p$$

式中脚标 e 表示弹性; 脚标 p 表示塑性。

根据 Mises 提出的塑性位势理论, 塑性流动的方向 (塑性应变增量矢量的方向) 与塑性位势函数 Q 的梯度方向相一致, 即

$$d\epsilon_p = d\lambda \frac{\partial Q}{\partial \sigma}$$

其中 dλ 是一正的有限量, 它的具体数值和材料硬化法则有关; Q 是塑性位势函数, 一般说它是应力状态和塑性应变的函数; 对于任一应力分量, 有

$$d\epsilon_{ij} = d\lambda \frac{\partial Q}{\partial \sigma_{ij}}$$

式中  $\sigma_{ij}$  —— 应力张量分量。

其中 dλ 的一般表达式为

$$d\lambda = \frac{(\partial f / \partial \sigma_{ij}) D_{ijkl}^e d\epsilon_{kl}}{(\partial f / \partial \sigma_{ij}) D_{ijkl}^e (\partial f / \partial \sigma_{kl}) + (4/9) \sigma_e^2 E^p}$$

应力应变的增量关系式

$$d\sigma_{ij} = D_{ijkl}^{ep} d\epsilon_{kl}$$

其中

$$D_{ijkl}^{ep} = D_{ijkl}^e - D_{ijkl}^p$$

$D_{ijkl}^p$  称为塑性矩阵, 它的一般表达式是

$$D_{ijkl}^p = \frac{D_{ijmn}^e (\partial f / \partial \sigma_{mn}) D_{rskl}^e (\partial f / \partial \sigma_{rs})}{(\partial f / \partial \sigma_{ij}) D_{ijkl}^e (\partial f / \partial \sigma_{kl}) + (4/9) \sigma_e^2 E^p}$$

对于九维应力空间, dλ 和  $D_{ijkl}^p$  可以简化为

$$d\lambda = \frac{(\partial f / \partial \sigma_{ij}) d\epsilon_{ij}}{(2/9) (\sigma_e^2 / G) (3G + E^p)}$$

$$D_{ijkl}^p = \frac{(\partial f / \partial \sigma_{ij}) (\partial f / \partial \sigma_{kl})}{(1/9) (\sigma_e^2 / G^2) (3G + E^p)}$$

式中 f —— 结构载荷。

各种硬化材料在九维应力空间中的具体表达式如下:

对于各向同性硬化材料

$$d\lambda = \frac{s_{ij} d\epsilon_{ij}}{(2/9) (\sigma_e^2 / G) (3G + E^p)}$$

$$D_{ijkl}^p = \frac{s_{ij} s_{kl}}{(1/9) (\sigma_e^2 / G^2) (3G + E^p)}$$

式中  $s_{ij} = \sigma_{ij} - \sigma_m \delta_{ij}$  —— 偏斜应力张量分量;

$$\sigma_m = \frac{1}{3} (\sigma_{11} + \sigma_{22} + \sigma_{33}) \text{ 是平均应力;}$$

$\delta_{ij}$  —— Kronecker delta。

对于理想塑性材料

$$d\lambda = \frac{s_{ij} d\epsilon_{ij}}{2\sigma_{e0}^2 / 3}$$

$$D_{ijkl}^p = \frac{s_{ij} s_{kl}}{\sigma_{e0}^2 / 3G}$$

式中  $\sigma_{e0}$  —— 材料初始屈服应力。

对于运动硬化材料

$$d\lambda = \frac{(s_{ij} - \bar{a}_{ij}) d\epsilon_{ij}}{(2\sigma_{e0}^2 / 9G) (3G + E^p)}$$

$$D_{ijkl}^p = \frac{(s_{ij} - \bar{a}_{ij}) (s_{kl} - \bar{a}_{kl})}{(\sigma_{e0}^2 / 9G^2) (3G + E^p)}$$

对于混合硬化材料

$$d\lambda = \frac{(s_{ij} - \bar{a}_{ij}) d\epsilon_{ij}}{[2\sigma_e^2 (\epsilon^p, M) / 9G] (3G + E^p)}$$

$$D_{ijkl}^p = \frac{(s_{ij} - \bar{a}_{ij}) (s_{kl} - \bar{a}_{kl})}{[2\sigma_e^2 (\epsilon^p, M) / 9G] (3G + E^p)}$$

式中  $\bar{a}_{ij}$  —— 移动张量的偏斜分量

将以上应力应变关系的一般表达式改写成矩阵形势为

$$d\lambda = \frac{(\partial f / \partial \sigma)^T D_e d\epsilon}{(\partial f / \partial \sigma)^T D_e (\partial f / \partial \sigma) + (4/9) \sigma_e^2 E^p}$$

$$D_p = \frac{D_e (\partial f / \partial \sigma) (\partial f / \partial \sigma)^T D_e}{(\partial f / \partial \sigma)^T D_e (\partial f / \partial \sigma) + (4/9) \sigma_e^2 E^p}$$

### 2 弹塑性增量分析有限元格式

基于增量形式虚位移原理有限元表达格式的建立步骤和一般全量形式的完全相同。首先将各单元内的位移增量表示成节点位移增量的插值形式

$$\Delta u = N \Delta \delta^e$$

再利用几何关系, 得到

$$\Delta \epsilon = B \Delta \delta^e$$

再由增量形式的最小势能原理, 并由虚位移的任意性, 得到有限元系统的平衡方程

$${}^t K_{ep} \Delta \delta = \Delta P$$

其中  ${}^t K_{ep}$ ,  $\Delta \delta$ ,  $\Delta Q$  分别是系统的弹塑性刚度矩阵, 增

量位移向量和不平衡力向量。它们分别由单元的各个对应量集成, 即

$$\left. \begin{aligned} {}^e K_{cp} &= \sum_e {}^e K_{cp}, \Delta \delta = \sum_e \Delta \delta^e \\ \Delta P &= {}^{+ \Delta} P_i - {}^a P_i = \sum_e {}^e P_i - \sum_e {}^e P_i^e \end{aligned} \right\}$$

并且

$$\begin{aligned} {}^e K_{cp} &= \int_{V_e} N^{Tr} D_{cp} B dV \\ {}^{+ \Delta} P_i^e &= \int_{V_e} N^{Tr+ \Delta} \bar{F} dV + \int_{S_e} N^{Tr+ \Delta} \bar{T} dS \\ {}^e P_i^e &= \int_{V_e} B^{Tr} \sigma dV \end{aligned}$$

上式中 ${}^{+ \Delta} P_i^e, {}^e P_i^e$ 分别代表外加载荷向量和内力向量, 所以 $\Delta P$ 称不平衡力向量, 如果 ${}^e P_i^e$ 和 ${}^e P_i$ 满足平衡的要求, 则 $\Delta P$ 表示载荷增量向量。

### 3 非线性方程组的解法

非线性问题的有限元离散化的结果将得到下列形式的代数方程组

$$K(\delta)\delta = P \quad \text{或} \quad \Psi(\delta) = K(\delta)\delta + f = 0$$

其中 $f = -P$ 。在以节点位移作为未知量的有限元分析中, 一次施加全部载荷, 然后逐步调整位移, 使上式得到满足。

#### 3.1 直接迭代法

对于方程

$$K(\delta)\delta + f = 0$$

假设有某个初始的试探解:

$$\delta = \delta_0$$

代入上式的 $K(\delta)$ 中, 可以求得被改进了的一次近似

解

$$\delta = -(K^0)^{-1} f$$

其中

$$K^0 = K(\delta_0)$$

重复上述过程, 可以求得 $n$ 次近似解

$$\delta^n = -(K^{n-1})^{-1} f$$

一直到误差的某种范数小于某个规定的容许小量 $\epsilon'$ , 即

$$\|e\| = \|\delta^n - \delta^{n-1}\| \leq \epsilon'$$

上述迭代过程可以终止。

#### 3.2 牛顿迭代法

对于方程

$$\Psi(\delta) = K(\delta)\delta + f = 0$$

设 $\delta_n$ 是上式的第 $n$ 次近似解, 一般地, 有

$$\Psi(\delta_n) = K(\delta_n)\delta_n + f \neq 0$$

由迭代式

$$\delta_{n+1} = \delta_n - [K_n^r]^{-1} \Psi_n$$

式中 $K_n^r = \frac{d\Psi}{d\delta}$ ——切线刚度矩阵。

重复上式, 直至满足精度要求为止。

#### 3.3 修正牛顿法

修正牛顿法就是将牛顿法中每次迭代中切线刚度矩阵求逆的过程省略, 将第一次迭代时的切线刚度矩阵 $K_0^r$ , 并求出逆矩阵, 用于每次迭代过程的切线刚度矩阵, 则迭代公式成为

$$\delta_{n+1} = \delta_n - [K_0^r]^{-1} \Psi_n$$

这样每一次迭代可以节省很多时间, 尤其是求解大型问题。



# 第 10 章 几何非线性问题的有限单元法

## 1 大变形情况下的应变和应力

### 1.1 应变的度量

在固定的笛卡尔坐标系内的一物体，在某种外力的作用下连续地改变其位形，如图 40.10-1 所示。用  ${}^0x_i$  ( $i=1, 2, 3$ ) 表示物体处于 0 时刻位形内的坐标，由于外力作用，在以后的  $t$  时刻，物体运动并变形到新的位形，用  ${}^t x_i$  ( $i=1, 2, 3$ ) 表示，其中左上标表示什么时刻物体的位形。物体的位形变化可以看作是从  ${}^0x_i$  到  ${}^t x_i$  的一种数学变换，即

$${}^t x_i = {}^t x_i({}^0x_1, {}^0x_2, {}^0x_3)$$

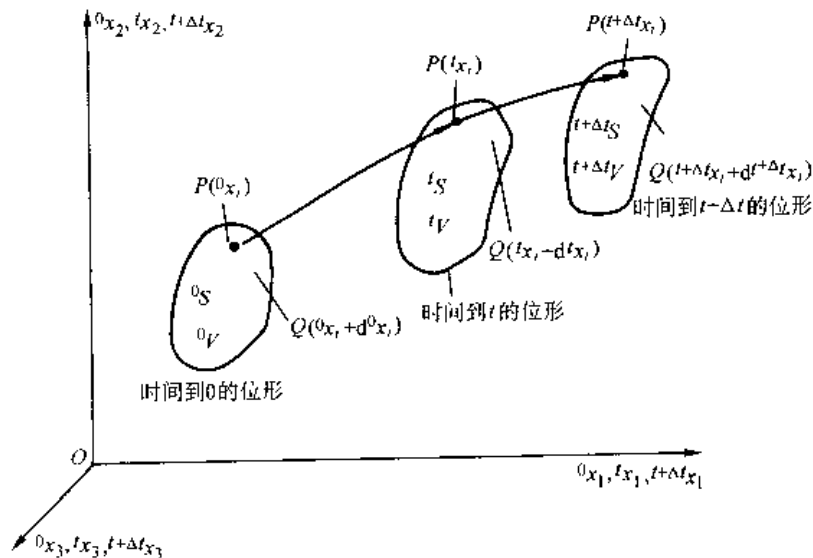


图 40.10-1 笛卡尔坐标系内物体的运动和变形

其中左上标表示该量对什么时刻位形的坐标求导数，右下标“ $\cdot$ ”后的符号表示该量对之求偏导数的坐标号。

利用以上各式可以得到变形的度量，即物体上任意两点之间的线段长度在变形前后的变化，对此有两种表示，即

$$\begin{aligned} ({}^t ds)^2 - ({}^0 ds)^2 &= ({}^0 x_{k,\cdot} {}^t x_{k,\cdot} - \delta_{ij}) d^0 x_i d^0 x_j \\ &= 2 {}^0 \epsilon_{ij} d^0 x_i d^0 x_j \\ ({}^t ds)^2 - ({}^0 ds)^2 &= (\delta_{ij} - {}^t x_{k,\cdot} {}^0 x_{k,\cdot}) d^t x_i d^t x_j \\ &= 2 {}^t \epsilon_{ij} d^t x_i d^t x_j \end{aligned}$$

从而定义了两种应变张量，即

$$\begin{aligned} {}^0 \epsilon_{ij} &= \frac{1}{2} ({}^0 x_{k,\cdot} {}^t x_{k,\cdot} - \delta_{ij}) \\ {}^t \epsilon_{ij} &= \frac{1}{2} (\delta_{ij} - {}^t x_{k,\cdot} {}^0 x_{k,\cdot}) \end{aligned}$$

根据变形的连续性要求，这种变换必须是一一对应的，也即变换是单值连续的，同时上述变换应有唯一的逆变换，即

$${}^0 x_i = {}^t x_i({}^t x_1, {}^t x_2, {}^t x_3)$$

利用上述变换，可以将  $d^0 x_i$  和  $d^t x_i$  表示成

$$d^0 x_i = \left( \frac{\partial^0 x_i}{\partial^t x_j} \right) d^t x_j, d^t x_i = \left( \frac{\partial^t x_i}{\partial^0 x_j} \right) d^0 x_j$$

引用符号

$${}^0 \epsilon_{i,j} = \frac{\partial^0 x_i}{\partial^t x_j}, {}^t \epsilon_{i,j} = \frac{\partial^t x_i}{\partial^0 x_j}$$

则  $d^0 x_i$  和  $d^t x_i$  可表示成

$$d^0 x_i = {}^0 \epsilon_{i,j} d^t x_j, d^t x_i = {}^t \epsilon_{i,j} d^0 x_j$$

${}^0 \epsilon_{ij}$  称为 Green-Lagrange 应变张量 (简称 Green 应变张量)，它是用变形前坐标的，即它是 Lagrange 坐标的函数。 ${}^t \epsilon_{ij}$  称为 Almansi 应变张量，它是用变形后坐标表示的，即它是 Euler 坐标的函数。其中左上标表示用什么时刻位形的坐标表示的，即相对于什么位形度量的。

当位移很小时，上式中的位移导数的二次项相对于它的一次项可以忽略，这时 Green 应变张量  ${}^0 \epsilon_{ij}$  和 Almansi 应变张量  ${}^t \epsilon_{ij}$  都简化为小位移情况下的无限小应变张量  $\epsilon_{ij}$ ，它们之间的差别消失，即

$${}^0 \epsilon_{ij} = {}^t \epsilon_{ij} = \epsilon_{ij}$$

另外，在大变形条件下， $({}^t ds)^2 - ({}^0 ds)^2 = 0$  意味着  ${}^0 \epsilon_{ij} = 0$  和  ${}^t \epsilon_{ij} = 0$ ，反之亦然，即物体为刚体运动的必要而充分条件是  ${}^0 \epsilon_{ij}$  和  ${}^t \epsilon_{ij}$  的所有分量到处为零。

### 1.2 应力的度量

在大变形问题中，平衡方程和与之相等效的虚功原理是用从变形后的物体内截取出的微元体来建立的，如图 40.10-2 所示。在从变形后的物体内截取出的微元体上面定义的应力张量称为 Euler 应力张量，用  ${}^t\tau_{ij}$  表示。

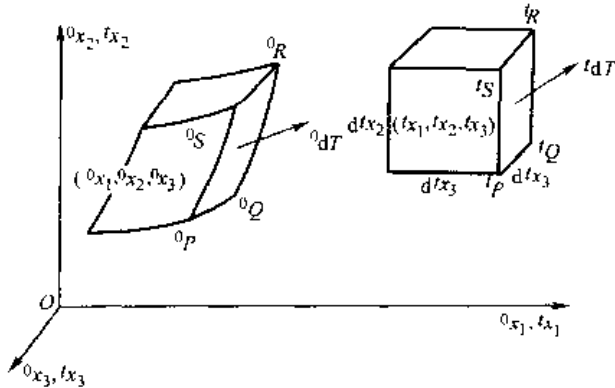


图 40.10-2 应力的度量

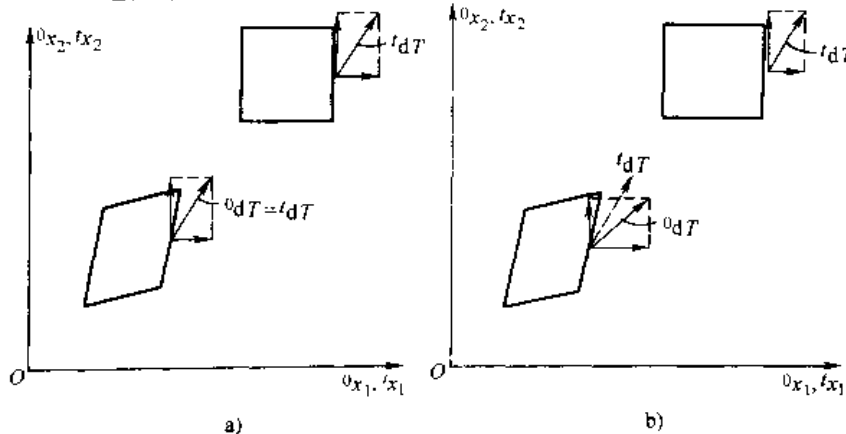


图 40.10-3 二维情况 Lagrange 和 Kirchhoff 应力规定的示意图

其中  ${}^0v_j$  是变形后面积微元  ${}^0dS$  上法线的方向余弦。类似上式所表示的关系用于变形前的位形，可得出以下两个关系式，如用 Lagrange 规定，则有

$${}^0dT_i^{(L)} = {}^0T_{ij} {}^0v_j {}^0dS = {}^0dT_i$$

如用 Kirchhoff 规定，则有

$${}^0dT_i^{(K)} = {}^0S_{ij} {}^0v_j {}^0dS = {}^0x_{i,j} {}^0dT_j$$

式中  ${}^0v_j$  —— 变形前面积微元  ${}^0dS$  上法线的方向余弦；

${}^0T_{ij}, {}^0S_{ij}$  —— 第一类和第二类 Piola-Kirchhoff 应力张量，或分别称为 Lagrange 应力张量和 Kirchhoff 应力张量。

左上标  $t$  —— 应力张量是属于变形后（时刻  $t$ ）位形的；

左下标  $0$  —— 变形前（时刻  $0$ ）位形内度量的。三种

在分析过程中，应变是用变形前坐标表示的 Green 应变张量，则需要定义关于变形前位形的应力张量与之对应。因此，假设变形后位形某一表面上的应力是  ${}^t dT / {}^t dS$ ，相应的变形前位形的该表面上的虚拟应力是  ${}^0 dT / {}^0 dS$ ，其中  ${}^0 dS$  和  ${}^t dS$  分别是变形前和变形后的面积微元。 ${}^0 dT$  和  ${}^t dT$  之间的相应关系通常有以下规定（参看图 40.10-3 所示）

1) Lagrange 规定：

$${}^0dT_i^{(L)} = {}^t dT_i$$

上式规定变形前面积微元上的内力分量和变形后面积微元上的内力分量相等。

2) Kirchhoff 规定：

$${}^0dT_i^{(K)} = {}^0x_{i,j} {}^t dT_j$$

上式规定  ${}^0dT^{(K)}$  和  ${}^t dT$  用和变换  $d^0x_i = {}^0x_{i,j} d^t x_j$  相同的规律相联系。

变形后位形的应力分量与内力的关系式如下

$${}^t dT_i = {}^t \tau_{ij} {}^t v_j {}^t dS$$

应力张量之间的变换形式如下

$${}^t \tau_{ij} = \frac{{}^t \rho_t}{{}^0 \rho_0} x_{i,k} {}^0 T_{kj} = \frac{{}^t \rho_t}{{}^0 \rho_0} x_{i,\alpha} {}^0 x_{j,\beta} {}^0 S_{\beta\alpha}$$

$${}^0 T_{ij} = {}^0 S_{i\alpha} {}^0 x_{\alpha,j}$$

式中  ${}^0 \rho, {}^t \rho$  —— 变形前位形和变形后位形的材料密度  
Lagrange 应力张量  ${}^0 T_{ij}$  是非对称的，不适合用于应力应变关系，而 Kirchhoff 应力张量  ${}^0 S_{ij}$  是对称的，更适用于应力应变关系。在小变形情况下，由于  ${}^0 x_{i,j} \approx \delta_{ij}$ ， ${}^0 \rho / {}^t \rho \approx 1$ ，这时可以忽略  ${}^0 S_{ij}$  和  ${}^t \tau_{ij}$  之间的差别，它们都蜕化为工程应力  $\sigma_{ij}$ 。

对于依赖于材料变形历史的非弹性问题，通常情况下需要采用增量理论进行分析，其中材料本构关系应采用微分型或速率型的。因此，在连续介质力学中还定义了一种其分量不随材料刚体转动而变化的速率型

的应力张量, 即 Jaumann 应力速率张量  $\dot{\sigma}'_{ij}$

$$\dot{\sigma}'_{ij} = \dot{\tau}'_{ij} - \tau'_{ip} \Omega'_{pj} - \tau'_{jp} \Omega'_{pi}$$

其中上标“·”表示对时间的导数,  $\Omega'_{ij}$  是旋转张量

$$\Omega'_{ij} = \frac{1}{2} \left( \frac{\partial' u_j}{\partial' x_i} - \frac{\partial' u_i}{\partial' x_j} \right) = \frac{1}{2} (u'_{j,i} - u'_{i,j})$$

它的物理意义是表示材料的角速度。

Jaumann 应力速率张量是对称张量, 它是不随材料微元的刚体旋转而发生变化的客观张量。与它对偶的应变速率张量是

$$\dot{\epsilon}'_{ij} = \frac{1}{2} \left( \frac{\partial' u_i}{\partial' x_j} + \frac{\partial' u_j}{\partial' x_i} \right) = \frac{1}{2} (u'_{i,j} + u'_{j,i})$$

$\dot{\epsilon}'_{ij}$  也是对称的, 且是不随材料微元的刚体旋转而发生变化的客观张量。 $\dot{\sigma}'_{ij}$  和  $\dot{\epsilon}'_{ij}$  在物理上分别代表真应力和真应变的瞬时变化率。

## 2 几何非线性问题的表达格式

在涉及几何非线性问题的有限单元法中, 通常采用增量分析的方法。考虑一个在笛卡尔坐标系内运动的物体, 增量分析的目的是确定此物体在一系列离散的时间点  $0, \Delta t, 2\Delta t, \dots$  处于平衡状态的位移、速度、应变、应力等运动学和静力学参量。

为得到用以求解时间  $t + \Delta t$  位形内各个未知变量的方程, 首先建立虚位移原理。与时间  $t + \Delta t$  位形内物体的平衡条件相等效的虚位移原理可表示为

$$\int_{t+\Delta t} \tau'_{ij} \delta_{i+\Delta t} \epsilon'_{ij} {}^{t+\Delta t} dV = {}^{t+\Delta t} Q$$

其中  $\delta_{i+\Delta t} \epsilon'_{ij}$  是相应的无穷小应变的变分。 ${}^{t+\Delta t} Q$  是时间  $t + \Delta t$  位形的外载荷的虚功。因为上式所参考的时间  $t + \Delta t$  位形是未知的, 所以方程不能直接求解。在实际分析中采用以下两种格式进行求解。

### (1) 全 Lagrange 格式 (T.L. 格式)

在此格式中, 与时间  $t + \Delta t$  位形内物体的平衡条件相等效的虚位移原理被转换为参考物体初始 (时间  $0$ ) 位形的等效形式, 也即方程中所有变量都是以初始位形为参考位形, 方程表示为

$$\int_{t_0} {}^0 S_{ij} \delta {}^0 \epsilon_{ij} {}^0 dV = {}^0 Q$$

式中  ${}^0 S_{ij}$  和  ${}^0 \epsilon_{ij}$  分别是时间  $t$  到  $t + \Delta t$  位形的 kirchhoff 应力和 Green 应变的增量, 并都是参考于初始位形度量的。

进一步考虑, Green 应变  ${}^0 \epsilon_{ij}$  可以表示成

$${}^0 \epsilon_{ij} = {}^0 e_{ij} + {}^0 \eta_{ij}$$

式中  ${}^0 e_{ij} + {}^0 \eta_{ij}$  —— 关于位移增量  $u_i$  的线性项和二次项 (非线性项)。

$${}^0 e_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i} + {}^0 u_{k,j} {}^0 u_{k,i} + {}^0 u_{k,i} {}^0 u_{k,j})$$

$${}^0 \eta_{ij} = \frac{1}{2} u_{k,i} u_{k,j}$$

### (2) 更新的 Lagrange 格式 (U.L. 格式)

在此格式中, 与时间  $t + \Delta t$  位形内物体的平衡条件相等效的虚位移原理被转换为参考物体更新 (时间  $t$ ) 位形的等效形式, 也即方程中所有变量都是以时间  $t$  位形为参考位形, 方程可表示为

$$\int_{t_0} {}^{t+\Delta t} S_{ij} \delta {}^{t+\Delta t} \epsilon_{ij} {}^{t+\Delta t} dV = {}^{t+\Delta t} Q$$

其中  ${}^{t+\Delta t} S_{ij}$  和  ${}^{t+\Delta t} \epsilon_{ij}$  分别是时间  $t + \Delta t$  位形的 kirchhoff 应力张量和 Green 应变张量, 它们都是参考于时间  $t$  位形。

与全 Lagrange 格式类似, 更新 Lagrange 格式的 Green 应变可以分解为

$${}^t \epsilon_{ij} = {}^t e_{ij} + {}^t \eta_{ij}$$

其中

$${}^t e_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i}) + {}^t \eta_{ij} - \frac{1}{2} u_{k,i} u_{k,j}$$

## 3 大变形条件下的本构关系

在实际分析中, 从结构变形特点考虑, 可以将大变形问题进一步区分为两类问题: 大位移、大转动、小应变问题和大位移、大转动、大应变问题 (简称为大应变问题)。从材料特点考虑, 实际问题又可以区分为弹性问题和非弹性问题。前者应力和应变之间有一一对应的关系, 而不依赖变形的历史。后者则应力和应变之间不存在一一对应的关系, 而与变形的历史有关。

### 3.1 弹性

#### 3.1.1 大位移、大转动、小应变情况

对此情况采用小变形线弹性本构关系的推广形式得

$${}^0 S_{ij} = {}^0 D_{ijkl} {}^0 \epsilon_{kl}$$

式中  ${}^0 S_{ij}$  —— Kirchhoff 应力张量;

${}^0 \epsilon_{kl}$  —— Green 应变张量;

${}^0 D_{ijkl}$  —— 常数弹性本构张量。

对于三维应力状态

$${}^0 D_{ijkl} = D_{ijkl} = 2G \left( \delta_{ik} \delta_{jl} + \frac{\mu}{1-2\mu} \delta_{ij} \delta_{kl} \right)$$

式中  $G, \mu$  —— 材料弹性常数。

#### 3.1.2 大应变情况

对于大应变情况, 在连续介质力学中用超弹性来表征这种材料特性。此时假定材料有一应变能函数  ${}^0 W$ , 它是 Green 应变张量  ${}^0 \epsilon_{kl}$  的解析函数, 但不限于  ${}^0 W = \frac{1}{2} \int_0^1 D_{ijkl} {}^0 \epsilon_{ij} {}^0 \epsilon_{kl}$  的形式, 它可能包含  ${}^0 \epsilon_{kl}$  的高次项。

从 \$W\$ 导出 Kirchhoff 应力张量 \$S\_{ij}\$, 即

$$S_{ij} = \rho \frac{\partial W}{\partial \epsilon_{ij}}$$

由此可以得到切线本构张量

$$D_{ijkl} = \rho \frac{\partial^2 W}{\partial \epsilon_{ij} \partial \epsilon_{kl}}$$

在实际分析中, 采用更新的 Lagrange 格式时, 联系 Euler 应力张量 \$\tau\_{ij}\$ 和 Almansi 应变张量 \$\epsilon\_{kl}\$ 的本构关系式是

$$\tau_{ij} = D_{ijkl} \epsilon_{kl}$$

其中

$$D_{ijkl} = \frac{\rho^t}{\rho_0} x_{i,m} x_{j,n} D_{mnpq} x_{k,p} x_{l,q}$$

### 3.2 非弹性

#### 3.2.1 大位移、大转动、小应变情况

因为 Kirchhoff 应力张量 \$S\_{ij}\$ 和 Green 应变张量 \$\epsilon\_{ij}\$ 是不随材料微元的刚体转动而变化的客观张量, 并且在小应变情况下数值上就等于工程应力和工程应变, 因此利用它们建立现在情况的本构关系, 即

$$d_0 S_{ij} = D_{ijkl} d_0 \epsilon_{kl}$$

式中 \$d\_0 S\_{ij}\$, \$d\_0 \epsilon\_{kl}\$ —— Kirchhoff 应力张量和 Green 应变张量的微分;

\$D\_{ijkl}\$ —— 时间 \$t\$ 位形的、参考于时间 0 的切线本构张量, 它是 Kirchhoff 应力张量和 Green 应变张量的函数。

对于弹塑性变形情况 \$D\_{ijkl}\$ 和前一节材料非线性情况在形式上完全相同, 只是用 \$S\_{ij}\$ 和 \$\epsilon\_{ij}\$ 代替了其中的工程应力和工程应变。

#### 3.2.2 大变形 (包含大位移、大转动) 情况

在大应变情况下, \$S\_{ij}\$ 和 \$\epsilon\_{ij}\$ 在数值上不等于工程应力和工程应变, 不便于用来确定本构关系中的材料常数, 因此更便于应用真应力和真应变及其速率。Jaumann 应力速率张量 \$\dot{\tau}'\_{ij}\$ 和应变速率张量 \$\dot{\epsilon}'\_{ij}\$ 在物理上分别代表真实应力速率张量和真实应变速率张量, 在大应变情况下它们之间的本构关系可以表示为

$$\dot{\tau}'_{ij} = D'_{ijkl} \dot{\epsilon}'_{kl}$$

和前面小应变情况列出的格式相类似, 可以列出大应变情况的各个表达式, 只是其中的 \$S\_{ij}\$ 和 \$d\_0 \epsilon\_{ij}\$ 被 \$\tau\_{ij}\$ 和 \$d\_0 \epsilon\_{ij}\$ 所代替。

### 4 几何非线性问题的求解方法

用有限单元法求解几何非线性问题, 首先需要用

等参元对求解域进行离散, 每个单元内的坐标和位移可以用其节点值插值表示如下

$$x_i = \sum_{k=1}^n N_k^0 x_i^k, \quad x_i = \sum_{k=1}^n N_k^t x_i^k, \quad x_i = \sum_{k=1}^n N_k^{t+\Delta} x_i^k$$

$$u_i = \sum_{k=1}^n N_k^0 u_i^k, \quad u_i = \sum_{k=1}^n N_k^t u_i^k, \quad u_i = \sum_{k=1}^n N_k^{t+\Delta} u_i^k \quad (i = 1, 2, 3)$$

式中 \$x\_i^k\$ —— 节点 \$k\$ 在时间 \$t\$ 的 \$i\$ 方向坐标分量;

\$u\_i^k\$ —— 节点 \$k\$ 在时间 \$t\$ 的 \$i\$ 方向位移分量; 其他 \$x\_i^k, t-\Delta x\_i^k, u\_i^k\$ 的意义类似;

\$N\_k\$ —— 和节点 \$k\$ 相关联的插值函数;

\$n\$ —— 单元的节点数。

在只考虑一个单元的情况下, 利用全 Lagrange 格式和更新的 Lagrange 格式可以得到不同的有限元求解方程。

应用全 Lagrange 格式的矩阵求解方程是

$$({}^0 K_L + {}^0 K_{NL}) \delta = {}^{t+\Delta} Q - {}^0 F$$

式中 \$\delta\$ —— 节点位移向量。

$${}^0 K_L = \int_V {}^0 B_L^T D_0 B_L dV$$

$${}^0 K_{NL} = \int_V {}^0 B_{NL}^T S_0 B_{NL} dV$$

$${}^0 F = \int_V {}^0 B_L^T S_0 dV$$

式中 \${}^0 B\_L, {}^0 B\_{NL}\$ —— 线性应变 \$e\_{ij}\$ 和非线性应变 \$\eta\_{ij}\$ 与位移的转换矩阵;

\$D\_0\$ —— 材料本构矩阵;

\$S\_0, S\_0\$ —— 第二类 Piola-Kirchhoff 应力矩阵和向量。

所有这些矩阵和向量的元素都是对应于时间 \$t\$ 位形并参考于时间 0 位形确定的。

类似的, 应用更新的 Lagrange 格式的矩阵求解方程是

$$({}^t K_L + {}^t K_{NL}) \delta = {}^{t+\Delta} Q - {}^t F$$

其中

$${}^t K_L = \int_V {}^t B_L^T D_t B_L dV$$

$${}^t K_{NL} = \int_V {}^t B_{NL}^T \tau_t B_{NL} dV$$

$${}^t F = \int_V {}^t B_L^T \tau_t dV$$

式中 \${}^t B\_L, {}^t B\_{NL}\$ —— 线性应变 \$e\_{ij}\$ 和非线性应变 \$\eta\_{ij}\$ 与位移的转换矩阵;

\$D\_t\$ —— 材料本构矩阵;

\$\tau\_t, \tau\_t\$ —— Cauchy 应力矩阵和向量。

所有这些矩阵和向量的元素都是对应于时间 \$t\$ 位形, 并参考于同一位形确定的。

## 第 11 章 有限单元法算例

### 1 空间桁架问题

**算例** 平面桁架结构如图 40.11-1 所示, 各杆件的材料和横截面积均相同, 试求各个杆件的节点位移和单元内力。

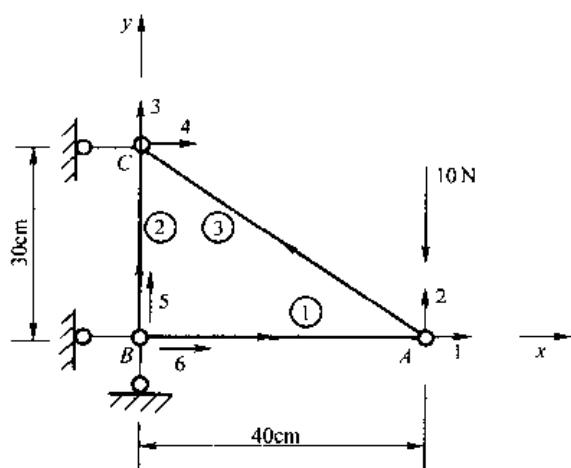


图 40.11-1 平面桁架结构

**条件:**  $E = 2.0 \times 10^5 \text{MPa}$ , 杆件横截面积  $A = 1.0 \text{cm}^2$

应用有限单元法求解, 将平面桁架结构离散化, 以每个杆为一个单元, 其整体坐标及各种编号如图 40.11-1 所示。每个节点有 2 个节点位移编号, 其中可动位移方向 (即自由度方向) 有 3 个。因此, 结构的未知节点位移阵为

$$\{\delta\} = [\delta_1 \quad \delta_2 \quad \delta_3]^T$$

每个单元杆中的箭头表出了局部坐标  $x$  轴的指向。

求解单元刚度矩阵, 首先求出各杆的方向角  $\alpha$ , 由第三节的坐标变换矩阵及相应公式可直接建立整体坐标下的单元刚度矩阵。

**单元 1** 因  $\alpha = 0$ , 故

$$k^{(1)} = \frac{EA}{l_1} \begin{matrix} & \begin{matrix} 6 & 5 & 1 & 2 \end{matrix} \\ \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ \dots & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{matrix} 6 \\ 5 \\ 1 \\ 2 \end{matrix} \end{matrix}$$

**单元 2**  $\alpha = 90^\circ$ ,  $c_x = \cos 90^\circ = 0$ ,  $c_y = \sin 90^\circ = 1$

$$k^{(2)} = \frac{EA}{l_2} \begin{matrix} & \begin{matrix} 6 & 5 & 4 & 3 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} & \begin{matrix} 6 \\ 5 \\ 4 \\ 3 \end{matrix} \end{matrix}$$

**单元 3**  $\alpha = 150^\circ$ ,  $c_x = -\frac{4}{5}$ ,  $c_y = \frac{3}{5}$

$$k^{(3)} = \frac{EA}{25l_3} \begin{matrix} & \begin{matrix} 1 & 2 & 4 & 3 \end{matrix} \\ \begin{bmatrix} 16 & 12 & -16 & -12 \\ 12 & 9 & -12 & -9 \\ -16 & -12 & 16 & 12 \\ -12 & -9 & 12 & 9 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 3 \end{matrix} \end{matrix}$$

以上各单元刚度矩阵的上边和右侧所标数字为单元节点位移方向在整体刚度矩阵中的行列位置, 由此整体刚度矩阵可通过各单元刚度矩阵叠加而得。又因对应 4、5、6 节点位移方向为零, 由划行划列法仅保留对应未知节点位移行列的矩阵元素, 可得:

$$K = EA \begin{bmatrix} \frac{1}{l_1} + \frac{16}{25l_3} & \frac{12}{25l_3} & -\frac{12}{25l_3} \\ \frac{12}{25l_3} & \frac{9}{25l_3} & -\frac{9}{25l_3} \\ -\frac{12}{25l_3} & -\frac{9}{25l_3} & \frac{1}{l_2} + \frac{9}{25l_3} \end{bmatrix} \\ = 2000 \begin{bmatrix} 378 & -96 & 96 \\ -96 & 72 & -72 \\ 96 & -72 & 405 \end{bmatrix}$$

节点载荷阵为

$$\{P\} = [0 \quad -10000 \quad 0]^T$$

求节点位移阵, 将  $K$  和  $\{P\}$  代入结构平衡方程, 解得

$$\{\delta\} = [-0.0267 \quad -0.12 \quad -0.015]^T \text{ (单位为 cm)}$$

计算单元内力, 因各单元只有轴向力, 可根据下式进行计算

$$N_i = \frac{EA}{l_i} [-c_x \quad -c_y \quad c_x \quad c_y] \{\delta'\}_i$$

**单元 1**  $\{\delta'\}_1 = [0 \quad 0 \quad \delta_1 \quad \delta_2]^T$

$$N_1 = \frac{EA}{l_1} [-1 \quad 0 \quad 1 \quad 0] \{\delta'\}_1 = -13.25 \text{kN}$$

**单元 2**  $\{\delta'\}_2 = [0 \quad 0 \quad 0 \quad \delta_3]^T$

$$N_2 = \frac{EA}{l_2} [0 \quad 1 \quad 0 \quad 1] \{\delta'\}_2 = -10.0 \text{kN}$$



### 3 轴对称壳体问题

球壳在均匀竖向载荷  $q$  作用下的变形和内力。

算例 计算如图 40.11-3 所示的周边铰支圆底扁

度  $t=0.12\text{m}$ , 球壳半径  $R=80\text{m}$ , 圆底半径  $a=25\text{m}$ 。  
条件:  $E=2.1 \times 10^6 \text{N/m}^2$ ,  $q=0.5 \text{N/m}^2$ , 球壳厚

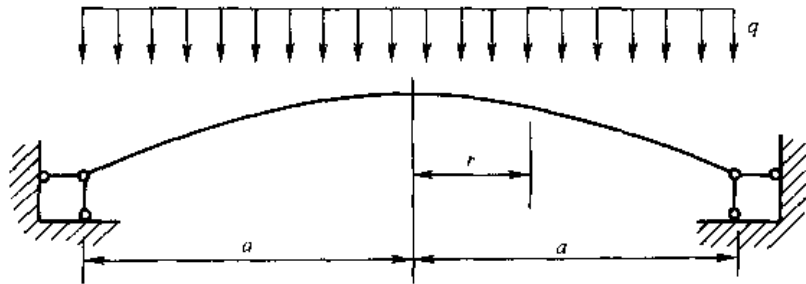


图 40.11-3 周边铰支圆底扁球壳

应用有限单元法求解,因其是对称问题,故采用截锥单元将扁球壳划分为八个单元,划分依据为底面极坐标,分别为  $r=4\text{m}$ ,  $r=7\text{m}$ ,  $r=10\text{m}$ , ...,  $r=25\text{m}$ 。边界条件为  $r=25\text{m}$  边铰支约束。

有限单元法计算出的竖向位移  $\delta$ 、经线弯矩  $M$ 、水平径向力  $H$ , 如表 40.11-1 所示。

表 40.11-1 计算结果

$r/\text{m}$	$\delta/ (10^{-2}\text{m})$	$M/ (\text{N} \cdot \text{m}/\text{m})$	$H/ (10^2\text{N}/\text{m})$
4	0.6651	$0.1076 \times 10^{-3}$	-0.2000
7	0.6648	$0.5261 \times 10^{-3}$	-0.2000
10	0.6631	$0.3436 \times 10^{-1}$	-0.1999
13	0.6625	$-0.5317 \times 10^{-2}$	-0.1998
16	0.6786	$-0.1112 \times 10^{-2}$	-0.2000
19	0.7158	$0.3465 \times 10^{-1}$	-0.2016
22	0.6158	0.1961	-0.2022
25	0.0000	0.0000	-0.1906

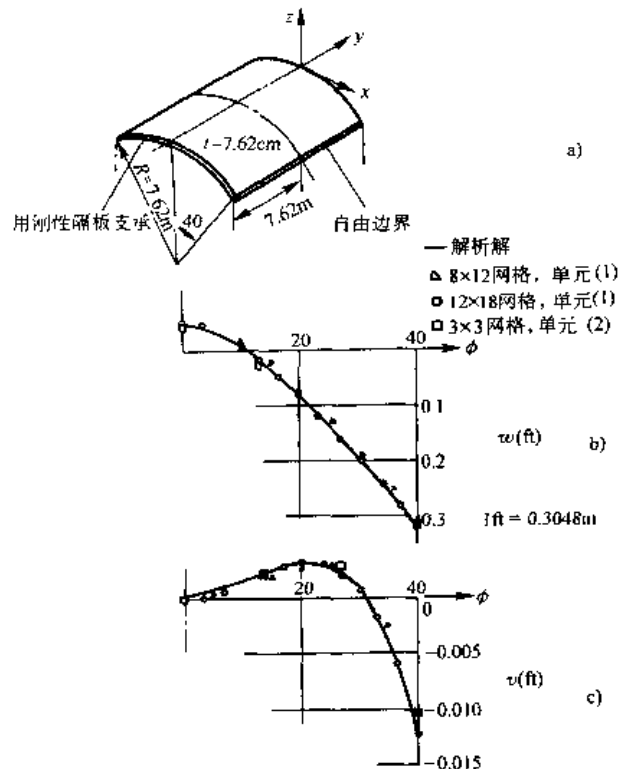


图 40.11-4

a) 圆柱壳屋顶 b) 自重作用下中央截面的垂直位移 c) 端部支承处的纵向位移

### 4 一般壳体问题

算例 如图 40.11-4a 所示为圆柱壳屋顶,两端由隔板支持,承受自重作用。计算圆柱壳屋顶在自重作用下中央截面的垂直位移和端部支承处的纵向位移。

条件: 几何尺寸如图所示,  $E=2.07 \times 10^4 \text{MPa}$ ,  $\mu=0$ 。

应用有限单元法求解,用折板代替曲面壳体进行计算。引用两组单元组合用不同的网格数目分别对屋顶进行离散化,以便比较它们的计算结果。

1) 平面应力单元是 3 节点三角形常应变单元、平板弯曲单元是 3 节点 9 个位移参数的非协调元。

2) 平面应力单元是 6 节点三角形二次元、平板弯曲单元是 12 个节点参数的三角形协调元,每个角节点参数是  $w$ ,  $\partial w/\partial x$ ,  $\partial w/\partial y$ , 边中节点参数  $\partial w/\partial n$ 。

边界条件为两个用刚性隔板支承边上  $u=0$ ,  $w=0$ , 两直边为自由边。

计算结果如图 40.11-4b、c 所示。从结果可见,当

网格划分比较精细时,这几种单元的结果和扁壳理论的解析结构比较,符合得相当好,同时也证明了用折板代替曲面壳体的可行性。

### 5 热传导问题

算例 如图 40.11-5 所示一受热载荷作用的厚壁圆筒,由对称性可取部分截面为研究对象,试分析其温度场。

条件: 材料在  $20^\circ\text{C}$  的热导率是  $7.5 \text{W}/(\text{m} \cdot \text{K})$ , 圆筒内壁温度  $500^\circ\text{C}$ , 外壁温度  $100^\circ\text{C}$ , 两端自由且绝热。内壁半径  $0.3\text{m}$ , 外壁半径  $0.5\text{m}$ 。

应用有限单元法求解,首先将圆筒截面离散化,划分为 75 个平面矩形单元,如图 40.11-6 所示。然后在分析网格上设定边界温度,即左边界(内壁)温度 500°C,右边界(外壁)温度 100°C。

有限单元法计算出的结果从左到右,温度成带状递减。底边六个节点(从左到右为:1、2、3、4、5、6)的温度值代表了整个截面上各温度带的温度值,如表 40.11-2 所示。

表 40.11-2 节点温度值

节点号	1	2	3	4	5	6
温度值	500.00	402.03	314.94	236.57	165.31	100.00

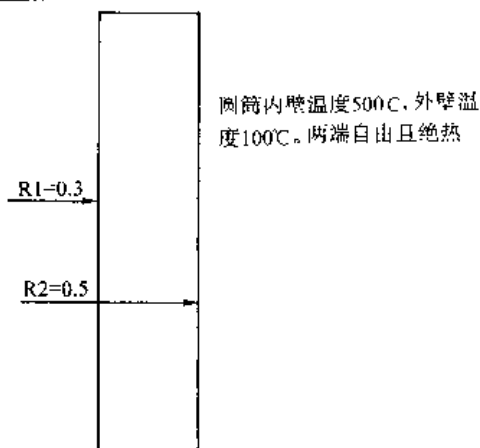


图 40.11-5 受热载荷作用的厚壁圆筒的计算分析模型(截面图)

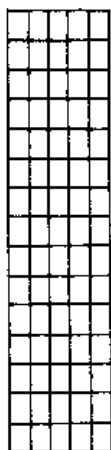


图 40.11-6 计算网格

### 6 动力学问题

**算例** 如图 40.11-7 所示,一正方形截面的刚梁长度为 10m,截面每边长 0.05m,在左端约束固定,试分析其自由振动的频率。

**条件:**  $L=10\text{m}$ ,  $b=h=5\times 10^{-2}$ ,  $E=3.0\times 10^{11}$ ,  $\mu=0.3$ ,  $\rho=7900\text{kg/m}^3$

应用有限单元法求解,首先将梁先分成单元 1、单

元 2 和节点①、节点②、节点③,如图 40.11-8 所示。

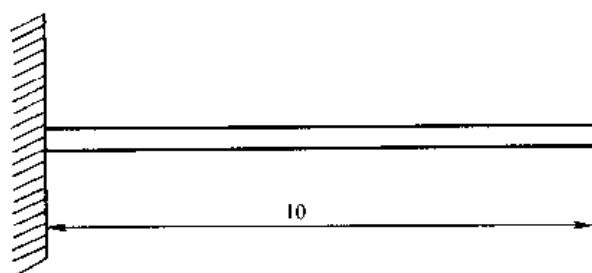


图 40.11-7 方形截面钢梁

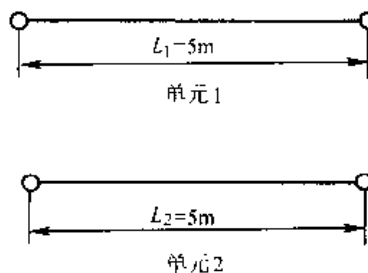


图 40.11-8 分解成两个单元

求质量矩阵,由第 8 章协调质量矩阵公式可得:

$$M = \rho A \int_0^l N^T N dx$$

将已知条件  $\rho=7900$  代入上式得到:

$$M = \begin{bmatrix} 36.6786 & 25.8631 & 12.6964 & -15.2827 \\ 25.8631 & 23.5119 & 15.2827 & 17.6339 \\ 12.6964 & 15.2827 & 36.6786 & -25.8631 \\ -15.2827 & -17.6339 & -25.8631 & 23.5119 \end{bmatrix}$$

求单元 1 的质量矩阵,由于节点①是固定端,所以可以将其忽略,为了计算方便,故简化成节点②节点③的质量矩阵,因单元 1 无节点③,故矩阵中对应节点③的元素都为零

$$M_1 = \begin{bmatrix} 36.6786 & -25.8631 & 0 & 0 \\ 25.8631 & 23.5119 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

求单元 2 的质量矩阵,由于单元 2 包含节点②和节点③,故其质量矩阵与  $M$  式相同,即

$$M_2 = \begin{bmatrix} 36.6786 & 25.8631 & 12.6964 & -15.2827 \\ 25.8631 & 23.5119 & 15.2827 & -17.6339 \\ 12.6964 & 15.2827 & 36.6786 & -25.8631 \\ -15.2827 & -17.6339 & -25.8631 & 23.5119 \end{bmatrix}$$

求总质量矩阵,将单元 1 和单元 2 的质量矩阵合并,即可得如下所示的总质量矩阵:

$$MM = M_1 + M_2 = \begin{bmatrix} 73.3572 & 0 & 12.6964 & -15.2827 \\ 0 & 47.0238 & 15.2827 & -17.6339 \\ 12.6964 & 15.2827 & 36.6786 & -25.8631 \\ -15.2827 & -17.6339 & -25.8631 & 23.5119 \end{bmatrix}$$



求刚度矩阵，由公式

$$K = EI \int_0^L B^T B dz$$

并将已知条件  $E = 3.0 \times 10^{11}$ ,  $I = \frac{bh^3}{12} = 5.2 \times 10^{-7}$

代入得到

$$K = \begin{bmatrix} 14976 & 37440 & 14976 & 37140 \\ 37440 & 124800 & -37440 & 62400 \\ -14976 & 37440 & 14976 & 37440 \\ 37440 & 62400 & 37440 & 124800 \end{bmatrix}$$

求单元 1 的刚度矩阵，与  $M_1$  的求法相似，得矩阵如下

$$K = \begin{bmatrix} 14976 & -37440 & 0 & 0 \\ -37440 & 124800 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

求单元 2 的刚度矩阵，同理由单元 2 包含节点 ②和节点 ③，故其刚度矩阵与  $K$  式相同，即

$$K_2 = \begin{bmatrix} 14976 & 37440 & -14976 & 37140 \\ 37440 & 124800 & -37440 & 62400 \\ -14976 & -37440 & 14976 & -37440 \\ 37440 & 62400 & -37440 & 124800 \end{bmatrix}$$

求总刚度矩阵，将单元 1 和单元 2 的刚度矩阵合并，即可得如下所示的总刚度矩阵：

$$KK = K_1 + K_2 = \begin{bmatrix} 29952 & 0 & -14976 & 37140 \\ 0 & 249600 & -37440 & 62400 \\ -14976 & -37440 & 14976 & -37140 \\ 37440 & 62400 & -37440 & 124800 \end{bmatrix}$$

求特征值，即  $\omega^2$ ，由第 8 章公式  $KK\phi - \omega^2 MM\phi = 0$  可解得特征值向量

$$\{\omega^2\} = \{37585.4, 4461.65, 390.034, 9.761133\}$$

又由公式  $f = \frac{\sqrt{\omega^2}}{2\pi}$  解得自由振动频率

$$\{f\} = \{30.8553, 10.6309, 3.1432, 0.497245\}$$

### 7 材料非线性问题

**算例** 轴向受到约束的承受内压作用厚壁圆筒，如图 40.11-9 所示，材料是弹性-理想塑性，并服从 V. Mises 屈服条件，试分析其变形和应力。

条件： $a = 1.0\text{cm}$ ,  $b = 2.0\text{cm}$ ,  $E = \frac{26}{3} \times 10^4 \text{N/mm}^2$ ,  $\mu = 0.3$ ,  $\sigma_s = 17.32 \text{N/mm}^2$ 。

用于计算的两种加载方案：

1) 内压按 0.5MPa 分级单调加载，直至塑性区达到厚度的 3/4。

2) 压力循环变化，具体是

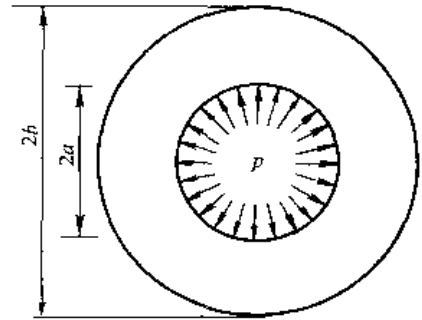


图 40.11-9

$p = 0.0 \rightarrow 10.0 \rightarrow 12.5 \rightarrow 0.0 \rightarrow 10.0 \rightarrow 12.5 \rightarrow 0.0 \text{MPa}$

应用有限单元法求解，在厚度方向用 4 个 8 节点轴对称单元建立有限元分析模型，如图 40.11-10 所示，刚度矩阵采用  $2 \times 2$  Gauss 积分。

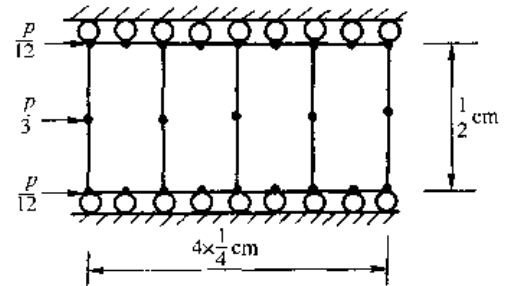


图 40.11-10

第一种加载方案的外表面径向位移如图 40.11-11 所示，用 Newton-Raphson 迭代和弹性常刚度迭代得到的结果都表示在图上。 $p = 12.5 \text{MPa}$  时（塑性区达到厚度的 1/2）的应力分布如图 40.11-12a、b 所示。

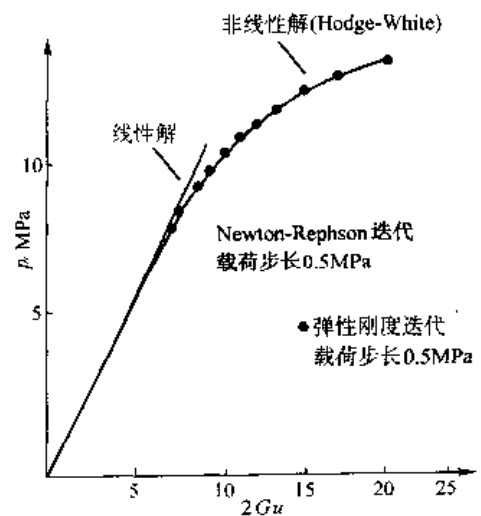


图 40.11-11 单调加载时的外表面径向位移

对于第二种加载方案，只采用了弹性常刚度迭代，外表面的径向位移如图 40.11-13 所示。

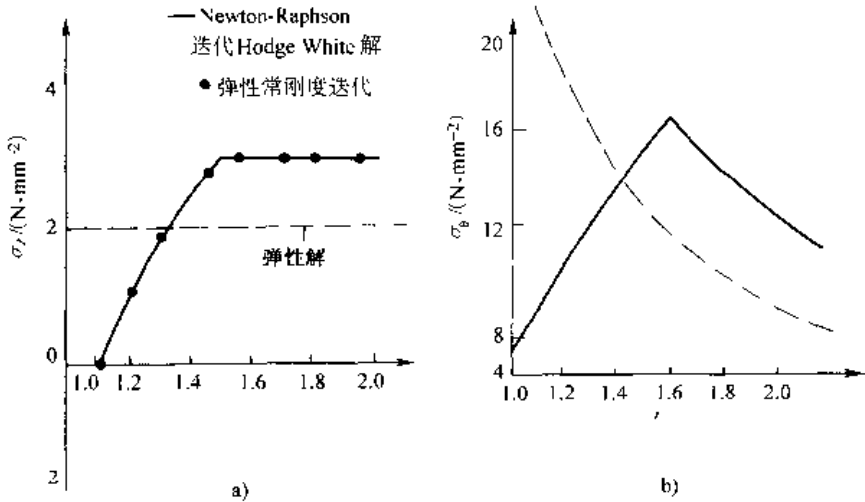


图 40.11.12

应用有限元法求解,首先对方板进行离散化。由于对称性,取出板的 1/4,用 4×4 矩形板单元对其划分。

引入边界条件,固定边的位移分量均为零,1/4 方板的内部边界只有法向位移,其余位移分量也均为零。然后用 Newton Raphson 法迭代求解,计算结果如图 40.11-14 所示。有限元法计算结果与莱维 (Levy) 理论解得接近,位移误差不到 1%。

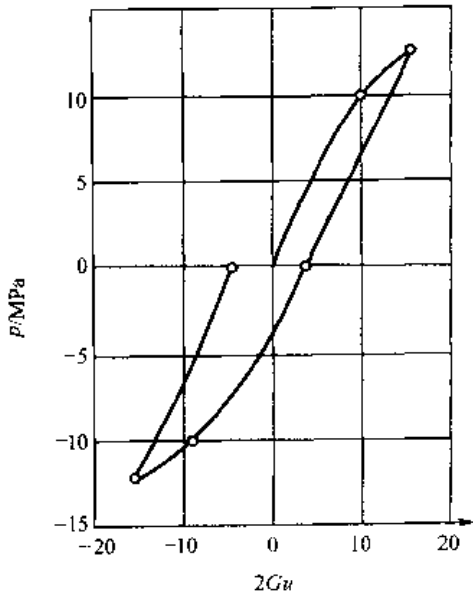


图 40.11-13 循环加载时的外表面径向位移

### 8 几何非线性问题

**算例** 均布载荷  $q$  作用下的简支方板,在板的中心内,沿边界的法向位移受到约束。试计算板中心点的挠度。

**条件:**板的平面尺寸为  $406\text{mm} \times 406\text{mm}$ ,厚度  $t = 2.54\text{mm}$ ,  $E = 207\text{GPa}$ ,  $\mu = 0.3$ 。横向载荷  $q = 103.5\text{kPa}$ ,分 6 级加荷,前 2 级荷载增量为  $\Delta q = 10.35\text{kPa}$ ,后 4 级荷载增量为  $\Delta q = 20.70\text{kPa}$ 。

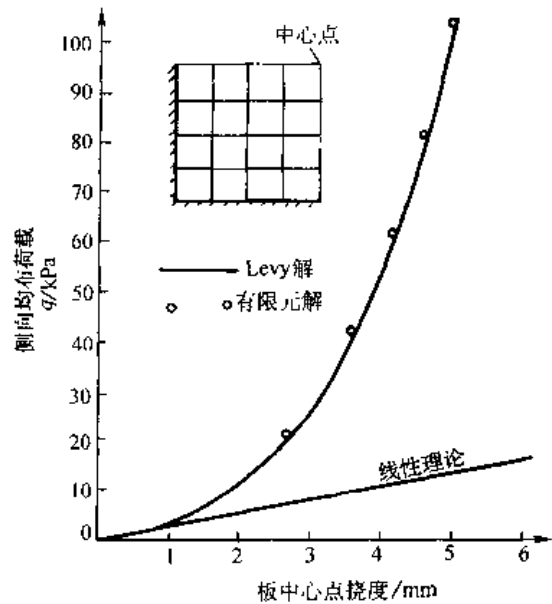


图 40.11.14 简支方板大挠度分析

## 第 12 章 典型有限元软件简介

### 1 MSC.NASTRAN

MSC.NASTRAN 是美国 The MSC Software Corporation (简称 MSC) 公司的产品, 1966 年, 美国国家宇航局 (NASA) 为了满足当时航空航天工业对结构分析的迫切要求主持开发大型应用有限元程序的招标, MSC 公司因一举中标而参与了整个 NASTRAN 的开发过程。到今天, MASTRAN 已成为各个工业领域一直认可和推崇的世界 CAE 工业最全面、应用最广泛的大型通用结构有限元分析软件, 也是国际合作和国际招标中工业工程分析和校验的有限元分析标准。

MSC.NASTRAN 的突出优势在于它具有极高的软件可靠性, 其计算结果已成为最高质量标准, 其输入输出格式也已成为当今 CAE 工业标准, 具有高度灵活的开放式结构和无限的解题能力。它不但容易使用而且具有十分强大的软件功能。通过不断地完善, 如增加新的单元类型和分析功能、提供更先进的用户界面和数据管理手段、进一步提高解题精度和矩阵运算效益等等。MSC.NASTRAN 不但适用于中小型项目, 对于处理大型工程问题也同样非常有效, 并且已成功地解决了超过 5 000 000 自由度以上的实际问题, 得到了世人的公认。

作为世界 CAE 工业标准及最流行的大型通用结构有限元分析软件, MSC.NASTRAN 的分析功能覆盖了绝大多数工程应用领域, 并为用户提供了方便的模块化功能选项, MSC.NASTRAN 的主要功能模块有: 基本分析模块 (含静力、模态、屈曲、热应力、流固耦合及数据库管理等)、动力学分析模块、热传导模块、非线性分析模块、设计灵敏度分析及优化模块、超单元分析模块、气动弹性分析模块、DMAP 用户开发工具模块及高级对称分析模块。

### 2 ANSYS

ANSYS 软件是融结构、流体、电场、磁场、声场分析于一体的大型通用有限元分析软件。由世界上最大的有限元分析软件公司之一的美国 ANSYS 开发, 它能与多数 CAD 软件接口, 实现数据的共享和交换, 如 Pro/Engineer, NASTRAN, Alogor, IDEAS, AutoCAD 等, 是现代产品设计中的高级 CAD 工具之一。

软件主要包括三个部分: 前处理模块, 分析计算模

块和后处理模块。前处理模块提供了一个强大的实体建模及网格划分工具, 用户可以方便地构造有限元模型; 分析计算模块包括结构分析 (可进行线性分析、非线性分析和高度非线性分析)、流体动力学分析、电磁场分析、声场分析、压电分析以及多物理场的耦合分析, 可模拟多种物理介质的相互作用, 具有灵敏度分析及优化分析能力; 后处理模块可将计算结果以彩色等值线显示、梯度显示、矢量显示、粒子流迹显示、立体切片显示、透明及半透明显示 (可看到结构内部) 等图形方式显示出来, 也可将计算结果以图表、曲线形式显示或输出。软件提供了 100 种以上的单元类型, 用来模拟工程中的各种结构和材料。

### 3 LS-DYNA

LS-DYNA 是世界上著名的通用显式动力分析程序, 能够模拟真实世界的各种复杂问题, 特别合适求解各种二维、三维非线性结构的高速碰撞、爆炸和金属成型等非线性动力冲击问题, 同时可以求解传热、流体及流固耦合问题。广泛应用于航空航天、汽车、国防工业、石油工业、核工业、电子工业、船舶工业等领域。

LS-DYNA 程序系列被公认为是显式有限元程序的鼻祖和理论先导, 是目前所有显式求解程序 (包括显式板成型程序) 的基础代码, 由 J. O. Hallquist 主持开发完成的。1988 年 J. O. Hallquist 创建 LSTC 公司, 推出 LS-DYNA 程序系列, 并于 1997 年将 LS-DYNA2D、LS-DYNA3D、LS-TOPAZ2D、LS-TOPAZ3D 等程序合成一个软件包, 称为 LS-DYNA。

LS-DYNA 程序版是功能齐全的几何非线性 (大位移、大转动和大应变)、材料非线性 (140 多种材料动态模型) 和接触非线性 (50 多种) 程序。它以 Lagrange 算法为主, 兼有 ALE 和 Euler 算法; 以显式求解为主, 兼有隐式求解功能; 以结构分析为主, 兼有热分析、流体-结构耦合功能; 以非线性动力分析为主, 兼有静力分析功能 (如动力分析前的预应力计算和薄板冲压成型后的回弹计算); 是军用和民用相结合的通用结构分析非线性有限元程序。

### 4 国产有限元软件

有限单元法在国内应用日益广泛, 国内高校和研究所等机构也开发出了具有自主知识产权的有限元软件, 并在科学研究和工程实际得到发展和应用。目前,

国内有限元分析软件主要有以下几种。

#### 4.1 GHFEA

GHFEA 是由华中理工大学、大连理工大学、清华大学于 1993 年至 1995 年联合开发,是国内第一套有自主知识产权的可与国外同类软件相比较的集成有限元建模、分析与结构优化及可视化系统。它支持有限元分析与优化设计的全过程,是工程技术人员进行工程结构分析设计的有力工具。GHFEA 主要面向结构的分析与设计,可广泛用于机械、航空、土木建筑、造船等行业。

GHFEA 由四个功能模块组成:

1) 有限元建模模块(FEM)。此模块包括几何模型的构造及简化操作,有限单元网格的生成,常用的杆、梁、板、壳、平面及实体等二十多种单元,并支持线性、非线性及用户定义模式等十多种材料,此外,系统还提供节点力、节点位移、节点温度、线分布压力、面分布压力以及位移边界条件的定义,同时,可进行模型的检查并提供 ADINA、SAPS 等有限元分析程序的接口。

2) 有限元分析模块(FEA)。有限元分析的功能主要有:静力分析、动力分析、屈伸稳定性分析、接触分析、弹塑性分析、几何非线性结构分析、热传导分析和高层建筑整体地震响应分析等。

3) 结构优化设计模块(FEA&SOP)。

4) 分析结果可视化模块(REVS)。此模块包括模型显示,数据分析、结果可视化、状态查询以及多项选择过滤等功能。

#### 4.2 有限元分析与结构优化微机软件系统 JFX95-W

有限元分析与结构优化微机软件系统 JFX-95W 是在多层子结构分析微机软件 JIGFEX-W、微机有限元分析软件 DDJ-W、计算机辅助结构优化设计软件 MCADS 的基础上发展的。它是一个通用的有限元分析和结构优化设计微机软件,适用于各种工程结构和部件的强度、刚度、稳定性、动力响应、热传导力学性能的分析计算和优化设计。其范围覆盖了航天、航空、机械、土木、建筑、车辆、水利、电力、石化等各个工业领域,是结构和产品设计研制及 CAD 系统中必要的分析计算软件工具。

JFX-95W 软件系统有如下子系统构成:结构静动力和热传导分析子系统、结构屈曲和几何非线性分析子系统、接触分析和弹塑性分析子系统、结构优化设计子系统、有限元建模前处理子系统、交互式图形后处理子系统。软件功能分为有限元分析、优化设计、前后

处理几个方面。

#### 4.3 有限元分析和优化设计系统 MAS

MAS 是由中国农机科学研究院研制的,能有效提高设计人员的设计能力和设计水平,快速获得合理、可靠、先进的设计方案和产品改进方案、有效降低成本、缩短设计周期并获得显著的经济效益。

MAS 系统由模型化前处理、有限元结构分析、优化设计、结果后处理等四大部分组成。

模型化前处理包括建模专家系统,交互式图形建模系统和模型评价系统三个部分。

有限元分析系统包括静力、动力分析系统、动态模拟系统、疲劳寿命分析、结构弹塑性分析、结构稳定性分析以及结构热分析。

优化设计系统包括结构灵敏度分析、结构静力优化和结构动力优化。

结果后处理系统包括图形后处理和数值后处理。

MAS 具有数值计算和智能技术相结合、分析和优化相结合、良好的开放性和良好的集成性的特点。

#### 4.4 线性与非线性结构分析程序/交互式有限元图形分析系统 APOLANS/INFEGAS

该程序由航空部 628 所与 631 所联合开发,以解决航空航天结构分析为主,同时具有广泛适应性的通用程序系统。该项程序还可用于核、船舶、汽车、压力容器、石油、化工、机械、民用工程、水工等各个工业领域的结构分析。其目的除解决工程中常见的一般性线性分析外,更着力于一系列非线性难度大的结构分析。强调程序对实际工程问题的实用性。

软件有六个独立的程序构成:结构分析前处理程序,解决面向用户的语言编译及自动分网;结构分析主程序;温度场计算的前处理程序;温度场计算程序;分析程序及显示程序的接口程序;图形显示与绘图程序。

#### 参 考 文 献

- 1 王勖成,邵敏.有限单元法基本原理和数值方法(第 2 版).北京:清华大学出版社,1997
- 2 朱伯芳.有限单元法原理与应用(第 2 版).北京:中国水利水电出版社,1998
- 3 张国瑞.有限元法.北京:机械工业出版社,1991
- 4 龙驭球.有限元法概论(第 2 版).北京:高等教育出版社,1991
- 5 夸克工作室.有限元分析基础篇 ANSYS 与 Mathematica.北京:清华大学出版社,2002

# 第41篇 虚拟设计

主 编 严隽琪

编写人 严隽琪

鲍劲松

# 第 1 章 虚拟设计总论

## 1 虚拟设计及其相关概念

虚拟设计是 20 世纪 80 年代提出的概念,20 世纪 90 年代得到重视和发展。虚拟设计的定义目前并不统一,虚拟设计和其相关概念,如虚拟制造、虚拟样机以及虚拟产品等容易混淆,有必要区分一下。

### 1.1 虚拟制造

虚拟设计和虚拟制造技术(Virtual Manufacture, VM)概念是相辅相成,不能分割(注:这里的制造不是指广义的大制造,只包括产品的设计,不包括生产)。虚拟制造是指利用计算机模型和仿真来实现产品的制造,它以信息技术、仿真技术、虚拟现实技术为支撑,在设计产品的物理实现之前,就能使人预测或感受到未来产品的形态、行为和性能,从而可以作出前瞻性的决策与优化实施方案。简而言之,是在计算机中实现的制造技术,它从根本上改变设计、试制、修改设计、规模生产的循环反复、串行式传统制造模式,是新一代制造模式。

随着 VM 这个新名词的出现,虽然,人们努力在给出 VM 的定义,但是往往都会从自身的研究领域或实践来定义它。因此这些已给出的定义总是会存在着缺点或不理想之处,甚至有些混乱。虚拟制造是一个处于发展中的新概念,它将随着技术的发展和人们科学研究的积累而不断得到丰富和深化。根据目前 VM 研究中出现的不同的侧重点,可以将 VM 归纳成不同的类别。

#### (1) 以设计为中心的虚拟制造

将制造信息加入到产品设计与工艺设计过程中,并且在计算机中进行数字化的“制造”,仿真多种制造方案,检验其可制造性或可装配性,预测产品性能和报价成本。其主要目的是通过不同逼真程度的“制造仿真”来优化产品的设计及其工艺过程,尽早发现设计中潜在的问题。

#### (2) 以生产为中心的虚拟制造

将仿真能力加入到生产计划模型中,其目的是方便和快捷地评价多种生产计划,是 IPPD(Integrated Product Process Design)到基于生产的物理层映射,检验新工艺流程的可信度,产品的生产效率,资源的需求状况(包括购置新设备、征询盟友等),从而优化制造环境的配置和生产的供给计划。

#### (3) 以控制为中心的虚拟制造

将仿真能力增加到(设备)控制模型中,提供对实际生产过程仿真的环境。其目的是在考虑车间控制行为的基础上,评估新的或改进的产品设计以及与车间生产相关的活动,从而优化制造过程,改进制造系统。

### 1.2 虚拟产品

虚拟产品(Virtual Product)也叫虚拟原型(Virtual Prototype)、虚拟样机、虚拟功能样机。虚拟产品是虚拟环境中的产品模型,是现实世界中的产品模型在虚拟环境中的映像。它可由多个通道与人发生交互作用。虚拟产品不仅包含了产品自身的定义数据,还要包含产品赖以“虚拟”的环境定义数据。虚拟产品开发是在虚拟环境中利用产品定义数据、环境定义数据以及产品与环境的相互作用规律来拟实反映产品在其生命周期中的演变历程。因此也可以说虚拟产品是基于知识的产品模型。

与一般的计算机仿真相比,虚拟环境中的产品分析不仅仅局限于对产品生命周期中各阶段分析评价的后处理或跟踪,而是能让用户从多通道中感受并驾驭产品演变过程,因而虚拟环境下的产品模型远远不能限于几何拓扑层次上建立的形体模型,也不是目前 CAD 软件提供的基于特征的参数化建模手段所能胜任的。其特点在于:

1) 虚拟产品是数字化定义的产品模型,在虚拟环境中表现出产品的几何和结构特性,并可与数字化的环境模型相作用。

2) 虚拟产品应该具有其物理原型的形体和表现,可以在计算机上逼真地展示产品性能。例如当用户关注产品的表面流场特性或曲面加工工艺时,虚拟产品的几何外形信息将是全面和完整的;当用户关注产品装配性能时,产品的结构信息将是详尽的。

3) 虚拟产品应该能以自然方式被人感受。如在产品外形方面,用户可触摸产品的几何和表面特性;在产品结构中,用户可拖拉零部件的位置;在产品的工作过程中,用户可感知运动件的作用范围和作用量的大小。

### 1.3 虚拟设计

虚拟设计就是以设计为中心的虚拟制造。必须认识到虚拟(Virtual),不等于虚幻、虚无,而是“本质”的

意思,是指通过数字化手段对物质世界的真实表现,亦即对真实世界的动态模拟,而这种数字化支撑手段主要包括虚拟现实技术、计算机仿真技术。设计则泛指围绕产品全生命周期的整个设计活动过程。因此可以这样认为,虚拟设计是虚拟现实技术在产品设计中的应用或实现。

上述虚拟产品是个相对于物理原型的概念,是具有一定功能的基于计算机的仿真系统。它不需构造物理原型,便能测试和评价多种设计的某些特征,达到缩短产品开发时间和降低成本的目的。如果虚拟原型是通过模拟工艺计划来构造的,那么这就是用虚拟设计来产生虚拟原型。所以虚拟原型的发展将增强虚拟设计的能力,反过来,虚拟设计为虚拟原型提供了应用环境。

#### 1.4 虚拟设计与其他概念之间的关系

任何新概念的出现都有其历史发展的必然性和鲜明的时代特征。虚拟设计也不是突然冒出来的,而是与在它出现之前出现的各种概念存在着千丝万缕的关系,但既然形成了一个公认的新的概念,就必然有其独到之处,以及在所有这些相关概念之外而独立存在的价值。从下面虚拟设计与仿真、计算机图形学、虚拟企业、可视化、多媒体技术等概念的比较中来深入理解虚拟设计的实质。

##### (1) 虚拟设计与仿真

虚拟设计依靠仿真技术来模拟设计、装配和生产过程,使设计者可以在计算机中“制造”产品。仿真是虚拟设计的基础,而虚拟设计是仿真的扩展。传统意义上的仿真一般不强调实时性,生成的可视化场景不会随用户的视点而变化、人与生成的场景没有自然方式交互,因此用户基本上是“旁观者”,没有“身临其境”的感觉。而在虚拟设计中,模型往往是动态的,因为虚拟现实使用户看到的景象会随视点的变化即时改变、可以和场景中物体进行交互,让感观接收到在真实情况中才能接收到的信息,增加了现场的沉浸感,使人“身临其境”。

##### (2) 虚拟设计与计算机图形学

虚拟设计依靠计算机图形学来建立计算机内的数字化模型,这种模型可以表达三维的立体数据,还可以像真实物品一样可视、可运动。从人机交互的自然程度来看,虚拟设计要远胜于在计算机图形学基础上建立的CAD系统,而且虚拟设计更强调用户感知方式的多样性,可使用户更“直接”地感知模型的物理特性。所以,CAD模型往往是提供给专业人员使用,而虚拟设计更支持不同技术背景,甚至非专业人员进行评价与讨论。从基础技术上看,计算机图形学又是虚拟设计的

基础,虚拟设计采用的虚拟现实技术,大量采用计算机图形技术来渲染场景,增加场景真实性。

##### (3) 虚拟设计与虚拟企业

虚拟企业是敏捷制造的动态组织形态。它是指为了赢得某一市场机遇,围绕某种新产品开发,通过选用不同组织或公司的优势资源,综合成单一的靠网络通信联系的阶段性经营实体,也称动态联盟。动态联盟具有集成性和实效性两大特点,它实质上是不同组织或企业间的动态集成,随市场机遇的存亡而聚散。在虚拟企业中的伙伴能共享生产、工艺和产品的信息,这些信息以数据的形式表示,能够分布到不同的计算环境中。虚拟企业和虚拟设计没有很强的相互依赖关系。虚拟企业强调网络环境下快速和敏捷的生产经营组织和管理,而虚拟设计的重点是仿真产品的生命周期中的各个活动。

##### (4) 虚拟设计与可视化

可视化是一种计算机方法,它将信号转换成图形或图像,使研究者能观察他们的模拟与计算,以丰富科学发现的过程。按照模拟执行与结果图像的结合程度,可将可视化研究分为三个层次:后处理、跟踪和控制。在后置处理中,图形显示是在数据计算后产生的,与数据源之间没有交互,其优点是结果图像可方便地重复显示,如计算流体力学、有限元计算结果的后置处理等等。跟踪是图形显示与计算过程同时进行,其特点之一是计算中间结果及最后结果都能及时显示,因而对计算中的错误可以及早发现,必要时可以停止执行,另一特点是图像直接从数据中产生,甚至数据无须写入存储介质中。控制(又称驾驭)是在计算过程中对参数进行修改,对数值模拟进行直接控制和引导。当可视化达到控制层次时,应可纳入VR的范畴,尽管可视化主要强调了“视觉”通道的人机界面。

虚拟设计为了揭露虚拟产品的本质——功能、性能,可视化技术是其关键的支撑技术。

##### (5) 虚拟设计与多媒体技术

多媒体技术中的“媒体”一词,其含义是信息的载体,如图形、文字、声音、图像等。计算机中的信息通过这些媒体来表达。如果媒体携带的信息种类仅有一种,如图像,则此媒体称为单媒体,如果媒体携带的信息是文字、声音和图像的综合,则此媒体称为多媒体。多媒体技术就是以计算机为核心的集图、文、声、像处理技术于一体的综合性处理技术。可见,多媒体技术虽然强调了多种信息“通道”,但主要还是提供视觉与听觉,感知范围没有虚拟设计中关键技术VR广,后者还包括了触觉、力觉等感知,而且一般说来,多媒体技术不强调人机交互性,如可视场景不随用户视点而变,因此,它提供的真实感不同于VR的存在感。

## 2 虚拟设计结构体系

面向产品开发的虚拟制造,又称虚拟产品开发,也可称为虚拟设计,以领域知识和拟实技术等关键技术为支撑,定义虚拟产品开发的“三要素”:产品定义数据、环境定义数据、产品与环境的交互作用规律;以揭示产品与其开发过程各个阶段特定的静、动态环境相互作用下产品特性的演变过程,通过高交互与高仿真的人机界面;以多种感知通道为设计者提供“自然”的感受与评价手段。

虚拟产品开发意味着用数字模型代替物理原型来进行产品设计中的分析与评价。它以产品的计算机辅助设计(CAD)模型为基础,应用不同的分析方法可以检验并改进设计结果。计算机辅助工程(CAE)系统,如有限元方法(FEM),动力学和运动学仿真可以提供有关产品性能的详细信息。将动画和虚拟现实技术与建模和仿真方法结合,即所谓的虚拟产品设计与开发(类似的词语如:虚拟/数字样机技术,数字化工程分析),为产品开发提供了新的可能方法,使不同的设计

方案可以进行快速评价。与物理原型相比,虚拟原型生成快,能直接操作和修改,且数据是可重用的。应用虚拟产品开发技术可以大大减少对真实原型的需求数量,并加快产品和工艺开发。这意味着极大地减少开发费用。

然而,新产品和工艺的开发涉及到来自不同学科的开发者。这就要求对所有相关人员(从技术开发者到市场专家)有一个综合的、合作的环境。如,开发人员要在产品开发的早期阶段对新产品进行测试,以缩短产品上市时间。虚拟产品设计与开发技术为设计师和工程师建立联接。实际上,虚拟产品开发项目在一个团队环境中执行可以获得很大成功,其中团队的各个小组分别工作于一个产品的不同部件,这些小组可以分布在同一栋大楼或是分散在世界各地。

图 41.1-1 为虚拟产品开发的体系框图。虚拟环境下的产品开发从概念设计、详细设计、数字样机支持下的各种性能分析和仿真,以及虚拟样机支持下的各种性能分析,在各种技术的支撑下,实现产品开发的各个阶段在计算机上的虚拟化。

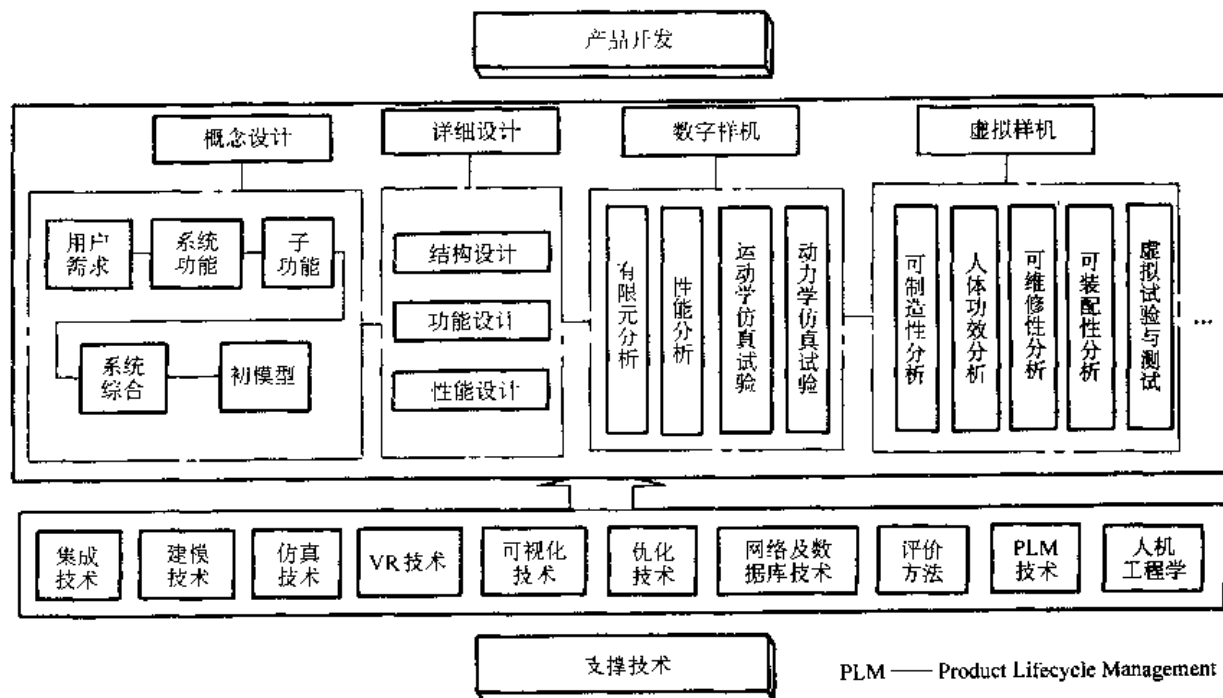


图 41.1-1 虚拟产品开发体系框图

产品开发过程就是组织进行产品开发的一系列步骤。目前有关产品开发过程的定义众多,以下几个定义颇具代表性:Feiler 和 Humphrey 给出的产品开发过程定义是“为达到产品的制造或改进目标而进行的一系列部分有序的步骤”,他们强调产品开发过程是一系列既定步骤;Osterweil 将产品开发过程上升为一种方法、路线,他定义产品开发过程为“创造一个产品的系统化方法”;Paul 等则定义产品开发过程为“人们用来

开发和维护产品的一些活动、方法、实践和信息转换”,这一定义将方法、实践等都列入了过程内容之中。另外对产品开发过程定义较有影响的是 MIT 的著名报告“90 年代的管理”中的对经营过程的定义:“是一系列相互影响的任务和功能,通过产生相应产品(Outcomes)来达到组织的经营成功”,产品开发过程显然是经营过程的一部分,这一定义强调组织过程的目的性(Purposeful nature of organizational process)。

PLM —— Product Lifecycle Management



以上这些定义都体现了人们对产品开发过程认识的逐步丰富,但是诸定义间也存在着差异和不足,为进一步讨论产品开发过程,定义产品开发过程如下:

产品开发过程是指从产品定义到产品批量生产之前这一段时间,它包括与产品开发有关的所有相关技术活动和管理活动。其中特定组织为满足某一需求利用各种资源使用工具和方法进行了创造性活动,过程的交付内容包括产品设计相关文档和数据。产品开发过程本身是一个将工程技术、方法、工具和人员集成并付诸产品开发实践的技术和管理框架。

该定义反映了人们对产品开发过程的一般认识,定义体现如下思想。

1) 产品开发过程主要涉及的是从产品定义到产品批量生产之前的这一段时间。

2) 产品开发过程涉及两个层面,即技术活动和管理活动,二者缺一不可。

3) 产品开发过程代表了既定组织进行产品开发的行为和方法,过程与特定的开发组织有密切的联系,组织文化、开发领域等对产品开发过程的设计和实施都有着重要的影响。

4) 产品开发是创造性活动,需要应用设计方法、技术和工具。

5) 产品开发过程本身是一个将工程技术、方法、工具和人员集成并付诸产品开发实践的技术和管理框架。这一观点代表了我们对产品开发过程工程研究与实践的最新认识,产品开发过程不单单意味着“属于某段时间”,也不单单是“一系列步骤”,也绝不仅仅“涉及到了一些方法和工具”,它是一个依附于既定经营领域、既定组织的开发技术和管理框架,其将工程技术、方法、工具和人力集成并付诸产品开发实践。

### 3 数字化辅助技术的发展趋势

在产品开发过程中应用计算机辅助工具及方法,实质上就是分别为了提高质量、减少时间和降低成本。市场竞争的激烈,要求技术能对不断增加的动态创新作出快速响应,使工程师不得不减少样机的设计和分析评价时间。图 41-1-2 所示为数字化辅助工程的发展趋势,可以看到,从二维/三维(2D/3D)的计算机辅助绘图工具,发展到三维(3D)的 CAD 系统(是一项产品开发中主要计算机辅助工具),它具有很多优势,但同时也有局限性,即对产品的感知不足。数字化模型(DMU, digital mock-up)主要体现在

在目前 CAE 工具的使用中,而虚拟环境下的数字化样机/虚拟制造则大大提高了基于计算机模型的可感知性,也是目前关注之所在。通过虚拟环境的应用,可实现两个目标:增强产品开发的自由度和获得更高层次的信息细节,从而更好地在产品开发的早期支持决策。而数字化工厂则更是高度集成了生产、控制、物流、成本、市场等信息,从工厂运作的角度进行虚拟产品开发。



图 41-1-2 数字化辅助工程的发展趋势

换言之,如图 41-1-3 所示,在产品开发的过程中(仅以产品设计到产品生产制造为内容),即使是采用 CAD 系统设计的产品,都必须通过样机的试制和实验,才能确定设计是否正确,产品性能和功能是否达到设计要求,设计中是否存在缺陷和问题。对于特殊要求的产品,必要时还需要进行原型机的破坏实验,是一种设计、试制、修改、再试制,直到成功的反复过程。用生产系统也可看成一种“产品”,它的规划设计,也存在一个不断规划、调试的过程。正如图 41-1-3 曲线所示,计算机技术在辅助产品设计和制造工艺设计、实际生产过程控制等中的应用水平已经较高,但在计算机辅助虚拟仿真实验的水平还相对较低,包括虚拟样机/虚拟制造在内的数字化技术发展,就是要不断地用基于物理的计算机仿真取代基于实物样机的实验,以真正实现较为完全的计算机上的数字化产品设计和生产制造。



图 41-1-3 从设计到制造内全局数字化方向发展

## 第 2 章 虚拟现实技术

### 1 虚拟现实的特点

虚拟现实技术来源于三维交互式图形学,目前已发展成为一门相对独立的学科,它与传统的计算机图形学和计算机仿真技术相比,有自己显著的特性:

#### (1) 沉浸感(IMMERSION)

虚拟现实系统中用户感知的环境是三维立体的,信息是多通道的,不仅能够获得视觉,还可以感受听觉、触觉等多种反馈信息,使得用户在虚拟环境中具有逼真的沉浸感。

在视觉上提供了立体显示模式,可以获得如同真实世界一样栩栩如生的三维视觉。

在声音反馈上,将用户置身于三维声音世界中,从而可以准确感受物体所在的方位以及虚拟环境中物体的交互作用的声音。

在触觉反馈上,用户在虚拟环境中触摸、抓取、移动物体时感受到物体的反作用力,甚至可以模仿真实世界中的受力情况。

在虚拟现实的沉浸感研究中,科学家正在研究动感、味觉甚至嗅觉。

#### (2) 交互性(INTERACTIVITY)

与目前基于 PWIM 模式,如二维菜单选项或命令输入等传统的交互方式极为不同,虚拟现实环境抛弃了鼠标的束缚,用户可以采用更为自然的三维操作或手势、语音等多通道信息来表达自己的意图,用户可采取现实生活中习以为常的方式来操纵虚拟环境中的物体,并改变其方位、属性或当前的运动状态,也可提供人在虚拟场景中的漫游。

#### (3) 想象空间(IMAGINATION)

在虚拟现实环境中,用户所见即所得,即用户处于多通道的三维空间中,从而可以充分发挥人的灵感和想象力,使设计的成功性和用户驾驭设计对象的能力得到空前的提高。因此,自然的人机交互方式、多种信息渠道、栩栩如生的三维空间,三者构成了用户与环境的统一。即通常所说的 3I(Immersion, Interaction, Imagination)。

#### (4) 实时性(Real-Time)

虚拟现实环境的最终目标是模拟真实的物理世界,因此虚拟现实系统应该可以按用户当前的视点位置和视线方向,实时地改变呈现在用户眼前的虚拟环境画面,并在用户耳边和手上实时产生符合当前场景

的听觉和其他各种感觉。正是这种实时性才使得用户产生极大的沉浸性,犹如身临其境。

### 2 虚拟现实系统组成及分类

#### 2.1 虚拟现实系统的组成

图 4.1.2-1 所示是一个典型的虚拟现实系统结构框图。头盔(或者其他产生立体视觉的装置),手套和话筒等输入设备为计算机提供输入信号,虚拟现实软件接到输入信号后,根据跟踪设备捕捉到的数据,调整当前的虚拟环境视图,对虚拟环境数据库作必要的更新,产生相应的视觉、听觉效果等。



图 4.1.2-1 典型虚拟现实系统结构框图

#### 2.2 虚拟现实系统的分类

虚拟现实系统按照不同的标准可以有多种分类方法,如按沉浸程度分为非沉浸式、部分沉浸式和完全沉浸式虚拟现实系统;按沉浸方式分为视觉沉浸、触觉沉浸、听觉沉浸和身体沉浸等;按用户参与的规模可分为单用户式、集中多用户式 and 大规模分布式系统等。目前使用较多的分类方法是既按沉浸程度又按用户规模

分类,大致分为:桌面虚拟现实系统、沉浸式虚拟现实系统和分布式虚拟现实系统。

#### (1) 桌面虚拟现实系统

该系统使用个人计算机和低端工作站来实现,采用三维立体眼镜,如液晶光闸眼镜,以及安装在计算机屏幕上方的发射器来同步系统产生立体图像,人佩戴立体眼镜后产生二维空间的视觉。

桌面虚拟现实系统价格比较低,可以配置各种虚拟现实设备,在各种专业应用中都具有潜力,但其沉浸感比较差。桌面虚拟现实系统目前大多采用 Windows/Linux 操作系统,配置一块能够输出立体图形的显卡,采用 OpenGL 就可以开发使用,见图 41-2-2。



图 41-2-2 桌面虚拟现实系统

#### (2) 沉浸式虚拟现实系统

利用头盔显示器和数据手套等交互设备把用户的视觉、听觉和其他感觉封闭起来,使参与者暂时与真实环境隔离,而真正成为虚拟现实系统内部的参与者,他/她可利用各种交互设备与虚拟环境交流。目前沉浸式虚拟现实系统主要有三种形式:沉浸桌 (ImmerseDesk)、全景墙 (PowerWall) 以及沉浸式 (CAVE)。

1) 沉浸桌 (ImmerseDesk) (见图 41-2-3) 沉浸桌比较适合用于装配、建筑领域,是适合 1~3 人合作设计的小型虚拟现实系统。



图 41-2-3 沉浸桌

2) 全景墙 (PowerWall) (见图 41-2-4) 有平面和弧形的两种,适合多用户参与、协同的设计。它明显的特点是具有非常宽阔的视野,在工业设计中用途最广,

可以演示 1:1 尺寸的虚拟样机模型。



图 41-2-4 全景墙

3) 沉浸式 (CAVE) (见图 41-2-5) CAVE 是 Computer Assisted Virtual Environment 的缩写,通常由 3,4 面墙组成,最复杂的系统有 5,6 面投影墙组成,通过投影设备将计算机分割的不同图像投射到不同的面上,产生立体图像。其适合大规模的数据采集、科学计算可视化,有很强的沉浸性,适合 3~5 人的科学研究。

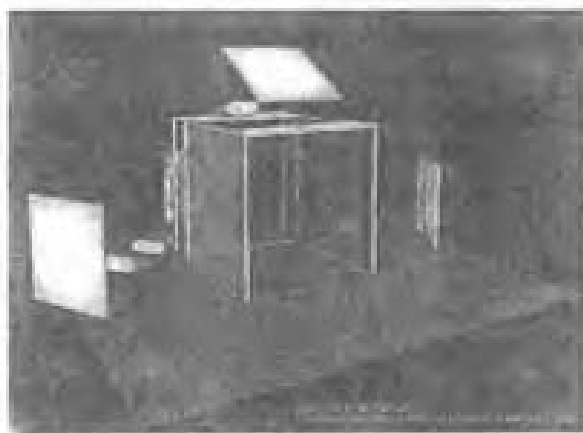


图 41-2-5 沉浸式

#### (3) 分布式虚拟现实系统

分布式虚拟现实系统 (DVR) 是一种基于网络的虚拟现实系统,它可使一组虚拟环境连成网络,使其能在虚拟域内交互,同时在交互过程中意识到彼此的存在。每个用户是虚拟环境中的一个化身 (Avatar)。它的基础是网络技术、实时图像压缩技术等,它的关键是分布交互仿真协议,必须保证各个用户在任意时刻的虚拟环境视图是一致的,而且协议还必须支持用户规模的伸缩性,常用的分布式协议是 DIS 和 HLA。

## 3 产生虚拟现实环境的工具集

### 3.1 产生虚拟现实环境的硬件

构建一个虚拟现实系统需要利用并集成高性能的计算机软硬件及各类先进的传感器,其硬件主要包括了能实现动态显示三维场景的计算机与图形加速设

备,实时跟踪和交互装置以及各类传感器组件等,如表 41-2-1、表 41-2-2 所示。

表 41-2-1 虚拟现实环境提供的各种感知刺激

感知刺激		范 围	VR 装置
视觉		感知可见光	图像生成系统,显示器
听觉		感知声波	声音合成器,音箱,喇叭
嗅觉		感知空气化学成分	气味传递装置
味觉		感知液体中化学成分	未实现
触觉	触觉	皮肤感知的触摸,温度,压力,纹理等	触觉传感器
	力觉	肌肉、关节、腱感知的力	力反馈装置
身体感觉		感知肢体、躯干位置	数据表
前庭感觉		平衡感知,由内耳感知头部的线加速度和角加速度	动感平台

表 41-2-2 虚拟现实环境检查各种动作的装置

用户动作	VR 装置
头部运动	头部跟踪装置
肢体、躯干运动	跟踪器,力反馈装置,空间球
手指运动	数据手套,控制装置,操纵杆,键盘
眼球运动	眼球跟踪器
语言	语音识别系统

### 3.1.1 虚拟现实系统的硬件组成

#### (1) 计算机与图像生成器

虚拟现实技术中最耗时的任务是图像的生成,图像生成器等硬件加速设备将大量图形操作封装在硬件中,如多边形生成、隐藏面剔除、消隐,甚至光照、光学追踪模型等。

主流的计算机及图形生成系统有:SGI Oxya Octane 等系列工作站,HP J.5V 系列工作站以及 Sun Blade 系列图形工作站,图形生成器主要有 SGI Infinite Reality Engine, NVIDIA GeForce, ATI 等高端图形生成器等。

#### (2) 交互设备

与虚拟世界进行交互的一个关键是跟踪真实世界物体位置并可自主操纵与控制的方式。虚拟现实环境主要的交互设备有:

1) 六自由度(6D)鼠标/跟踪球/游戏杆,见图 41-2-6。它们都带有额外的按钮和滑轮,不仅仅用于控制 XY 坐标变换,而且可以控制 Z 方向的旋转。全设备 3D 控制杆是一个 5D 游戏杆,它看起来象一个附在短棒上的网球,可以在左/右和上/下方向按钮这个球。



图 41-2-6

2) 数据手套,见图 41-2-7。数据手套可以在手指处装有传感器和全面位置/定位跟踪,数据手套其中大多数采用了基于指关节的光纤传感器和用于全位置的磁跟踪器,一般数据手套的操作通过手的姿势变化以及位置来和虚拟场景进行交互,还有一些带力反馈的数据手套,通过机械式,或小型充气泵等方式产生力觉,目前比较出名的是 VPL, Fake Space 的 PinchGlove, 5DT 公司出品的各种手套。

数据手套的位置精度由跟踪设备提供,姿态的多

少(即提供传感器 Sensor 多少)决定了数据手套的功能,目前 22 个 Sensor 的数据手套非常昂贵。

力反馈器,见图 41-2-8,通过机械设备产生力反馈,目前这类设备在眼柄上应用。

3) 魔棒(Wand),类似于三维鼠标,不过在魔棒上有多种按钮,可以控制多种功能 Wand 有很多种类,图 41-2-9 是 FakeSpace 出品的 NeoWand。

#### (3) 位置跟踪器

位置跟踪器提供位置(X, Y, Z)和三个定位测量(滚



图 41.2-7

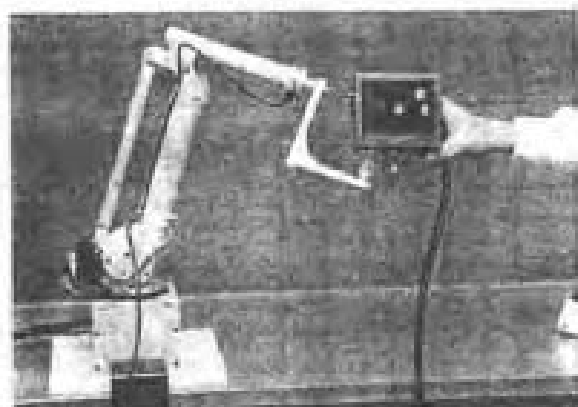


图 41.2-8



图 41.2-9

动角、俯仰角、航向角)。如图 41.2-10 为 Flock of Bird 跟踪系统,包括发射器、接收器以及各种传感器。位置跟踪的最大问题是反应时间,或在输入仿真引擎之前需要进行测量和数据处理的时间。通常的位置跟踪器有以下几种:

1) 机械传感器 可以提供快速和非常精确的跟踪。机械传感器的缺点是设备的笨重和移动的限制。其中 Exos 系统建立了用于手控制的类似骨架,并提供了力反馈。

2) 超声波传感器 能够进行位置跟踪和定位。它



图 41.2-10 跟踪系统。

具有一套发射器和接收器,发射器用于产生脉冲序列,相应的接收器能测量到延时信号。通过三角测量可以给出位置。超声波传感器的缺点是分辨率低,大延时和容易受到回声和环境中的其他噪声的干扰。

3) 磁跟踪器。使用线圈组来产生磁场。磁传感器决定磁场的强度和角度。这种传感器的限制是对于测量和处理有很长的反应时间,距离限制和受到磁场内磁性材料的干扰。然而,磁跟踪器仍不失为最好的方法之一。

4) 光学位置跟踪系统 近年已经发展起来,一种方法是使用栅格发光二极管和头盔式摄像机,发光二极管产生脉冲序列,摄像机图像被处理以决定闪光。这种方法的问题是空间受到限制(光栅限制)和缺少完整的移动(旋转)。另一种方法是使用许多视频摄像机来捕捉通过高速计算机修正的实时图像以跟踪对象,这里处理时间(高档计算机的花费)成为主要的限制因素。

5) 惯性跟踪器 一般仅提供转动的测量,且对于缓慢位置变化测量也不精确。目前已经发展起来。

#### (4) 立体视觉装置

1) 投影设备 虚拟现实的投影设备大多采用 CRT 三枪投影设备(图 41.2-11),目前高端的产品已经采用 DLP,亮度可以达到 5000 甚至更高。一般大型虚拟现实系统,如 CAVE Power-Wall 系统都采用多个投影设备,进行拼接而成,这种系统大多要求投影设备具有边缘融合功能(Edge Blending),目前主流的设备有 Barco 808s、909s,挪威 3D Perception 公司的最新产品 Image™ 166-2 边缘融合设备以及美国科视数字公司(Christie Digital)的 Mirage 系列,投影机如图 41.2-11 所示。

2) 立体视觉设备 立体视觉装置通过创建两幅不同的虚拟图像,每一幅对应一只眼睛,图像在两眼点之间进行等距离视点偏移计算,从而形成立体视觉。

① 立体眼镜 主要有两种技术来实现这两幅图像,一种称为被动式(Passive),这两幅图像可以并排放置,重叠在一起,图像能够通过置于眼前的不同的滤

波器投影到眼睛,立体图像使用偏正光眼镜或红/蓝玻璃眼镜,以提供天然的(无色彩)立体视觉,这类眼镜造价便宜,电影等娱乐业大量采用该类型。

一种称为主动式(Active),计算机交替产生左、右眼图像,使用液晶显示眼镜与显示器上的立体图像对,同步开关,遮挡左/右眼,使大脑很快收到交替图像,并融合图像为单一的场景和景深。所以需要高速的场频(最小为60Hz)来避免闪烁,否则用户佩戴后会感到头痛。这类眼镜较昂贵,主要有Crystal Eye公司的立体眼镜,见图41-2-12。

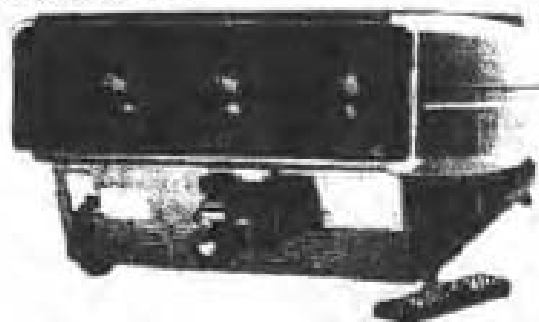


图 41-2-11 投影机



图 41-2-12

另外一种产生立体图像的方法是分屏的方法,如双目镜(BIOM)和头盔显示器(HMD)。

② 双目镜(BIOM) 双目镜是其中产生立体图像的方法是分屏的方法一种装置,见图41-2-13。它把屏幕分成两个部分同时显示左右眼图像。正对着显示器放置一种安在头部的观察器,帮助用户两个眼睛分离开,正确地看到立体的图像。

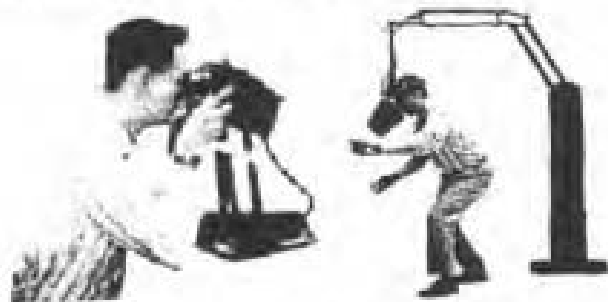


图 41-2-13

③ 头盔显示器(HMD) 虚拟现实系统中常用的一种硬件设备是头盔式显示器(HMD),是其中产生

立体图像的方法是分屏的方法另一种装置,见图41-2-14。它使用某种类型的头盔或护目镜,在每只眼睛前放置一个小型视觉显示器,并以特殊的光学技术来聚焦并延伸到可感觉到的视区范围。大多数HMD使用两个显示器,能够提供双目图像,也有的使用一个大显示器来提供高分辨率,但没有立体视觉。



图 41-2-14

大多数廉价的头盔显示器是使用LCD显示,其他的使用小型CRT显示器,就像是在手持式摄像机上用的那种。更贵一些的头盔显示器使用特别的沿头部架设的CRT显示器或光纤束从非头戴式显示器中输送图像。一个头盔显示器需要一个位置追踪器附加到头盔上,显示器也可以佩戴到身上以支撑和追踪。

### 3.1.2 可获得的硬件系统

#### (1) 图形工作站

SGI公司是目前VR硬件设备的核心提供商,他们主要提供高端图形工作站以及超级计算机,出名的有Onyx2、Onyx3000系列产品,它是当今最强大的可视化主机,专门为可视化计算而设计,其中包括纹理映射,体绘制和反走样显示,Clip Texture等。

HP公司J系列图形工作站也用来作为VR环境的图形生成系统。基于Windows的PC图形工作站由于其造价低廉,应用开发方便,得到普遍重视,各种PC图形工作站可以满足一定的VR应用。

图形卡:NVIDIA,ATI芯片系列高端显卡。

#### (2) 传感器

Ascension、Polhemus公司的各类电磁传感器是目前虚拟现实(VR)市场中最先进和最通用的传感器,通过采用低频磁场传感器和DSP技术可减小传感器延迟(达4ms),提高采样频率(120Hz),增大工作范围[最大到15ft(1ft=0.3048m)],常用的型号有Ascension公司的Flock of Birds,以及Polhemus的ISOTRAK I,FASTRAK等。

#### (3) 数据手套

Fake Space 5DT 公司的 Pinch Glove 以及 5DT 公司的 data glove, 其中有提供 5 和 16 个感觉的数据手套。

#### (4) 头盔

Virtual Research System 公司的 V6、V8 等头盔显示器(HMD)产品由于采用了有源 LCD 矩阵技术, 使其分辨率得到了改善。其中, V6 分辨率为  $640 \times 480$  像素, V8 为  $(640 \times 3) \times 480$  像素, 后者图像质量可与 CRT 头盔显示器媲美。而基于 CRT 的头盔显示器分辨率可达  $2048 \times 1152$ , 但其视角没有 LCD 显示器大, 且价格昂贵。

#### (5) 显示系统

Barco 公司的投影系统是 VR 显示系统的主要提供商, 有 CRT、LED 以及 DLP 系列投影系统, 投影屏幕的范围可以大到数 10m 范围。美国科视数字公司(Christie Digital)也提供类似投影系统。

### 3.2 虚拟现实的软件子系统

从系统角度来看, 虚拟现实软件必须充分管理和利用好各种计算资源、接口设备和系统资源, 提供给虚拟现实应用开发人员一个高性能的接口。用户接口方式可有以下几种:

1) 编译库 也称为 API 或 Toolkit 方式。这种接口方式最为灵活, 但要求开发人员必须是程序设计人员, 而且, 每次修改程序时都需要重新编译。它比较适合于开发复杂特定的虚拟现实应用系统, 如 OpenGL, Performer 等。

2) 描述语言 使用描述语言开发应用系统相对简单, 开发者无需有程序设计的知识背景就可以在较短的时间内学会使用描述语言。

3) 图形化 图形化界面使用户通过简单的点击即可开发应用系统。优点是无需程序设计背景知识而且使用方便直观, 缺点是难以开发复杂的应用系统, 如 VEGA, Relax 等。

虚拟现实软件所提供的环境一般包括对象编辑器、世界编辑器和集成运行环境。对象编辑器主要实现在对象坐标系中为物体几何建模和行为建模的功

能, 世界编辑器主要实现对象在世界坐标系中的行为建模和集成, 对象编辑器和世界编辑器有机地结合起来以实现虚拟环境的建模和管理, 集成运行环境提供了虚拟环境的实时表现场所, 它可以是一个复杂的分布式仿真环境。如图 41.2-15 所示是虚拟现实软件环境的框图。

#### 3.2.1 虚拟现实软件系统的组成

VR 系统的最基本部分可以分为输入处理器、仿真处理器、绘制处理器和虚拟环境数据库(World Database)。所有的部分必须考虑到处理的时间要求。每个延迟都意味着降低实时的感觉。

##### (1) 输入处理器

VR 程序输入处理器控制用于把信息输入到计算机的设备。目前有很多种输入设备: 键盘、鼠标、跟踪球、游戏杆、3D & 6D 位置跟踪器(数据手套、数据棒、头跟踪器、数据衣等), 网络 VR 系统还应该增加网络接收设备。语音识别系统对于 VR 也是一个输入设备。一般来说 VR 输入设备应该尽量简单, 对象应该以最小的延迟从系统的其他部分得到坐标数据。一些位置传感器增加一些滤波和数据平滑处理, 一些手套系统增加手势识别。

##### (2) 仿真处理

VR 程序的内核是仿真系统, 可以处理交互、对象的动作、物理原型的仿真和测定虚拟场景状态。网络 VR 应用可以使多个仿真运行在不同的机器上, 每一个都是不同的时间步骤。

仿真引擎使用户根据输入到虚拟世界中的任务(例如, 碰撞检测、脚本等)进行输入, 同时也检测在虚拟世界中将要发生的动作。

##### (3) 绘制处理

VR 程序的绘制处理是用来创建真实感的程序, 可以把绘制处理分为视觉、听觉、触觉和其他传感系统。每一绘制器会从仿真处理器那里得到虚拟世界状态的描述或直接从虚拟世界数据库中得到。

1) 视觉绘制 视觉绘制是最普通的处理, 它在计算机图像和动画领域已经有了很长的历史。影响 VR 图像绘制的主要因素是帧生成速率。它需要每秒产生 20 或 20 以上的帧数, 因为 20 帧/s 是人类眼睛能看到连续和平滑图像的最低限度。

2) 听觉绘制 VR 系统通过包含听觉部件大大增加了效果。它可以是单、立体或 3D 声音。后者是非常难的命题, 由于大脑趋向于在头的内部定位声音, 结果不足以做到立体效果。3D 声音的研究已经表明人类头部和耳朵的形状影响 3D 声音识别。声音可以是转化其他信息的方式, 如表面粗糙度。在沙子中拖动虚拟手

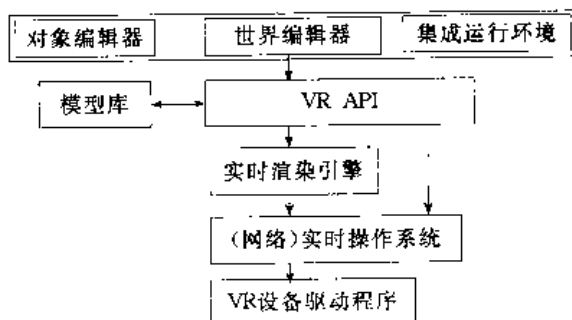


图 41.2-15 虚拟现实的软件环境

掌与在碎石中的声音是不同的。

3) 触觉绘制 几乎目前所有的系统都集中在力反馈和肌肉运动感觉上。这些系统能够相对于感觉给躯体提供很好的暗示。

4) 其他感觉 平衡和移动的感觉在 VR 系统中可通过移动平台以达到相当的精度。常用于飞行仿真和影院里提供一些运动暗示,这时人的注意力集成了其他的暗示来感知运动。

#### (4) 虚拟场景

虚拟场景本身需要被定义在一个“场景空间”里。作为计算机仿真的本质该场景需要限制。

1) 场景坐标 空间的一个主要约束是坐标即定位,需要在虚拟场景里在每一个对象的每个点上赋以数值。一般地,坐标代表笛卡尔坐标系的 X、Y 和 Z (长、宽、高)当然可以使用其他坐标系,如球坐标系,但是笛卡尔坐标是最常用的,而且坐标系之间的变换非常简单,一般程序实现都使用齐次坐标。

2) 场景的划分和环境的分离 通常一个虚拟场景很庞大,而由于硬件设备的限制,不能做到较好的实时性,经常把虚拟世界划分为多个场景,并提供在每种场景之间的变换。使得绘制可以计算更少的对象。

#### (5) 虚拟环境数据库

有关虚拟对象和场景数据是 VR 系统的一个重要部分。存储在场景数据库(场景描述文件)的最主要的对象就是代表场景的对象,描述对象行为或用户(发生在用户身上的)脚本,光线、程序控制和硬件设备支持。

1) 存储方法 场景信息的存储方法有很多,数据文件一般情况存为 ASCII 文本文件,但在很多系统里,用二进制文件代替 ASCII 文本文件,一些系统直接把场景信息编辑到应用程序里。最为出名的是 Open Flight 文件格式,它是 Multigen 公司虚拟场景数据库文件,描述了场景的存储等信息。

2) 对象 虚拟场景的对象可以是几何形状、层次性的对象组合、脚本或其他属性。对象的行为对系统的结构和设计有非常巨大的影响。属性列表通过名称来附加到对象(例如 cube.mass 表示立方体对象的大小),它们也可以是数量矢量或表达值,也可以附加在对象的脚本内部。

#### (6) 脚本和对象行为

仅仅由静态对象组成的虚拟世界几乎没有什么趣味。需要定义动作,当用户(或其他对象)与之交互时,对象执行自己的行为。可以将脚本分成三个基本的类型:动作脚本、触发脚本和连接脚本。

1) 动作脚本 动作脚本修改场景的位置、方向或者对象的属性、当前环境的光线或者摄像机。为简单化和提高速度,对于一个对象在任何一个距离仅仅一种

动作脚本被激活。动作脚本非常有效,其复杂度取决于我们允许脚本有多复杂。由于脚本的解释需要时间,而这影响帧数和延迟速率,所以必须注意。而且,脚本从层次树上绑定或拆开对象。

2) 触发脚本 触发脚本涉及当某些触发事件发生时,例如碰撞、接近或选择。VR 系统需要在每一次 Tick 评价触发器参数。对于接近检测器,它可以是从对象到三维视点的简单的距离检测或者是效果器(Effector)对象,碰撞检测是更为复杂的处理。

3) 连接脚本 连接脚本控制输入和输出设备到各种对象的连接。例如,连接脚本可以将手套设备连接到虚拟手对象。手套的移动和位置信息用来控制在虚拟世界里手对象的位置和行动。一些系统直接把函数集成到程序里,某些系统设计成 VR 系统几乎是一个连接脚本。

#### (7) 交互反馈

当虚拟光标选择或触摸对象时,用户必须给出某些交互反馈的指示。简单系统仅仅让用户看见光标(虚拟手)穿透对象,这样用户能够抓住物体或者选择物体,被选择的物体然后以某种方式被加亮。另外,产生声音信号可以表示碰撞。

#### (8) 硬件控制和连接

虚拟世界数据库可能包含关于硬件控制以及它们如何集成到应用程序中的信息。它们也有可能是程序代码的一部分。一些 VR 系统把这些信息放到配置文件中。硬件映射部分定义了输入/输出、数据速率和用于每个设备的其他参数。它也把设备的逻辑连接提供给虚拟世界的某些部分。

### 3.2.2 可获得的虚拟现实软件系统

1) EDS 软件公司 (<http://www.eds.com/products/plm/teamcenter/>) EAI 系列软件,包括 VisConcept、VisMockup、Jack 等。

2) Sense8 公司 (<http://www.sense8.com>)

World Toolkits——提供 C、C++ 库的软件包。

World Up——虚拟场景的建模工具,可以读取 dxf、3ds、nff 等格式文件。

World To World——网络环境下的虚拟现实软件。

3) Multigen-Paragim 公司 (<http://www.multigen-paragim.com>)

Multigen——强大的虚拟现实建模工具,支持多种文件格式。

Vega——视景仿真软件,提供强大的 C++、C 接口。

4) CG2 公司 VTree——CG2 产品包括



FACETS, VTree SDK, VTree Pro SDK 和 Mantis, 通过 VTree 的图形工具 Splice Tree 和 Audition 可以方便地实现视觉仿真、实时场景生成、娱乐冒险环境模拟、任务训练、事件重现等应用。

5) RealAx 公司 (<http://www.realax.com>) RealAX 系列软件——虚拟现实建模、仿真软件。

6) SGI 公司 (<http://www.sgi.com>) Performer —— 开发工具集, 可以实时渲染。

7) PTC 公司 (<http://www.ptc.com>) dVision——类似 EAI 的应用程序。

8) IBM 公司 (<http://www-306.ibm.com/software/applications/plm/enovia/>) ENOVIA DMU 解决方案可以使工程和工艺主管人员对任意复杂程度的数字样机和技术数据实现协同式工作, 包括高性能的可视化、浏览、审核、分析、仿真等。提供 Immersive VR 模块。

## 第3章 基于虚拟现实技术的新一代CAD技术

### 1 基于虚拟现实的CAD的特点

当今的CAD系统已发展到以实体造型为主的年代,比如CATIA、UG、PRO/E以及I-DEAS等软件,但就用户界面而言,无一例外地遵循WIMP(Windows-Icons-Menu-Pointer)的操作方式,其系统采用二维显示方式,设计者与系统的交互依赖于键盘、鼠标等设备。这种输入以串行性和精确性为特征,设计者每次只能利用一种输入设备来指定一个或一系列完全确定的指令或参数。三维的设计常常不得不分解为二维、甚至一维的操作,这使得三维设计过程异常复杂、乏味,约束了交互的效率。人们一直期望一种具备自然的交互方式的全新的CAD技术。直到最近几年,计算机技术的飞速发展,一种基于虚拟现实技术的CAD技术问世,使人们的期望逐步变成现实。

基于虚拟现实技术的新一代CAD(简称VR-CAD),采用的输入设备是:数据手套、三维鼠标、语音系统以及其他跟踪设备。它是基于多媒体的、高交互的、浸入式或半浸入式的三维计算机辅助设计环境。在这样的虚拟环境中,设计者不仅能够自始至终在三维空间里观察漫游和分析设计结果,而且能够直接在三维空间中通过三维操作、语音指令、手势等高度交互的方式进行三维实体建模和装配建模,并且最终生成精确的几何模型以支持详细设计与变动设计。同时能在同一环境中进行一些相关分析,从而满足工程设计和应用的需要。

VR-CAD还可以实时地提供各种反馈信息,包括视觉反馈、声音反馈、力反馈等,让设计者与被设计对象间建立双向的关系。

新一代的基于虚拟现实技术的CAD技术,其主要特点是:

#### (1) 自然的人机交互方式

1) 灵活的导航 为了让用户在不同的方位观察整个物体,对物体的每一个细节进行观察,还可以深入到物体的内部,查看物体内部细节,精确地选取物体,采用手势、语音、身体其他部位的运动跟踪等导航方式。

2) 逼真的视觉 VR-CAD中的模型不同于传统的CAD模型,渲染方式也不同于以前,它更加考虑到人在系统中的感受。因此首先对几何模型进行三角化剖分,然后再基于左右视点计算出一对图像以供立

体显示,再通过纹理、光照、渲染场景,让用户身临其境。

3) 三维、多通道的信息反馈 VR-CAD中,实时的三维立体声音反馈不仅有助于实时地了解设计过程,而且也是定位物体的一个非常有效的方法之一;设计过程力的反馈将调动设计者的触觉感官,更逼真地接近现实环境。

#### (2) 实时的约束识别和求解机制

满足物体精确定位支持详细设计,现有的虚拟设备,因为精度问题,不能使物体完成精确定位,为此,必须进行实时的约束识别和快速求解的研究。保证物体的设计、装配、形状分析、功能分析得以进行,支持详细设计。

#### (3) 适当的几何约束模型

虚拟环境的三维属性和高逼真性,给设计者更好的沉浸感,带来观察、分析模型的巨大优势。但对几何约束模型也提出了非常苛刻的要求,如支持多媒体的信息、识别手势、跟踪身体运动、反馈声音和力等,这些要求必须要包容在几何约束模型中。

## 2 VR-CAD的几何建模技术

### 2.1 传统的几何建模技术简介

进行设计,需要在计算机内用产品模型来表达设计的物体,人们最常用的表示方法是CSG表示和B-reps表示,而更经常的是将这两种方法合并在一起运用。

实体的B-rep表示模型能够表达出多面体各几何元素之间完整的几何和拓扑关系,并且允许对这种几何和拓扑关系进行修改。其中,体、面、边和顶点是最基本的几何元素。

另一种建模方法称为结构实体几何(Constructive Solid Geometry,或CSG),由两个物体间的并、交、差等欧拉操作生成一个新的物体。CSG应用基于三维物体的基本元素,如长方体、棱锥、柱体、锥体、球或者封闭的样条曲面。首先旋转两个基本元素,放置到空间某个位置。然后选择组合两个基本元素的体操作(并、交、差)。这时就形成了一个新的物体。

B-reps和CSG在实际应用中都有一定的欠缺,很难应用在一个复杂物体的建模中,融合两种建模技术为一体的混合建模是目前最主要的建模方式。

## 2.2 VR—CAD 中的几何建模

VR—CAD 中的几何建模指利用自然交互方式在虚拟场景中进行三维建模的过程,VR—CAD 中的几何约束模型是基于 CSG/B-reps 混合表示的一个扩展,它由 B-reps 表示、CSG 表示和几何约束表示三部分组成。然而目前 VR—CAD 中的几何建模还在发展阶段,还达不到商品化三维软件那样针对任意复杂三维模型的建模,作为一种未来三维建模技术,这里简单介绍 VR—CAD 中简单模型的几何建模方法。

### (1) VR—CAD 中的 B-reps 表示

1) 对于点的表示,采用了世界坐标系和局部坐标系下的双重坐标系表示。目的主要有两个:其一,使得对基本体素形状的调整能通过一种快速的修改操作完成,而无需重新生成该基本体素的边界表示;其二,有利于变动的传播和回溯。

2) 扩展面、边、点的表示属性,VR 操作中经常需要对实体模型进行选择操作,因此增加“选取点”属性,可以在三维空间中方便对面、边、点的选取,其中面、边的选取点设置在其几何中心,点的选取为其自身。

3) 增加面、边、点的父节点与其子节点的联系指针,利于在导航或者变动设计时定位与回溯。

4) 增加节点与临近相关点的联系指针,便于引入人力反馈、面音源、体音源等功能。

### (2) VR—CAD 中的 CSG 表示

为支持直接三维操作、实体显示以及显示定形、定位约束和变动设计,可在基本体素的属性中加入形状控制点(SCP)、定位方式(LP)、以及辅助元素(AE)等,而且 CSG 树的每个节点都当作一个独立物体,显式存

储其完整的边界表示,即显式地保存了它的顶点表、边表、面表。对于其中的叶节点而言,它们作为原始拓扑元素,在设计过程中始终不变。可以让几何约束总是施加在基本元素的原始拓扑元素上。

设计产品的过程是一个构造性的过程,通过实例化一个基本体素并将其附加到以前生成的相对简单的物体上以形成更为复杂的物体。在这个过程中,尽管有可能出现复杂的布尔运算,但绝大部分操作其实还是粘结(Glue)运算。此时,当基本体素的形状改变时,只要这种变化不影响到其父节点的拓扑结构,它的面、边、点的几何变化即可通过父指针逐级往上传递,达到目标物体。这样为快速变动传播提供了可能。

### (3) 基本体素的处理

对于 VR—CAD 系统而言,如何规范设计者与几何模型的交互方式以更有利于几何模型的定形定位,是一个不可回避的问题。

1) 形状控制点 为方便设计者在三维空间中调整基本体素的形状,可以引入“形状控制点”,【参考浙江大学,万华根论文】,形状控制点是基本体素一个特殊的点,通过它可以确定基本体素形状参数的空间点,用户可以方便的拖动该点改变基本体素的形状。因此实际操作中要考虑基本体素的形状控制点的可见性与可及性。从而无需通过键盘输入数值,就能方便地更改其形状。比如球体,形状控制点可以设在球心,而一个圆柱体,控制点设在底面与母线的交点上等。这些控制点的设定可以参考目前成熟的三维 CAD 商品软件。基本体素的形状参数通过形状控制点和有关映射自动生成。其过程可以概括如图 41.3-1 所示。

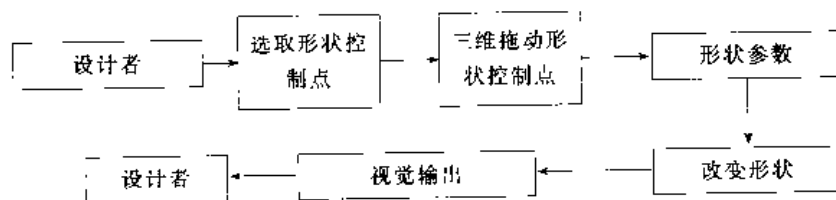


图 41.3-1 形状控制点操作过程

形状控制点分为两类:主形状控制点和从形状控制点。其中主形状控制点只有一个,它用来控制基本体素的总体形状,如图 41.3-2 中的小圆点所示。而从形状控制点如果存在,一般会有多个,他们用于微调基本体素的局部形状,并可能直接地影响到基本体素的形状参数,如图中的小叉号所示。

2) 定位方式 其次设定基本体素的定位方式,具体地说,定位方式指的是借助于若干特定种类的几何约束,通过方位已经确定的基本体素来决定方位待定的基本体素的方向、位置的方式。定位方式保证建模的

准确性以及虚拟造型交互过程的实时性。基本体素的定位方式很多种,总结起来分为以下几类:

#### 1 类定位方式

此定位方式是通过一个“面—面”贴合约束和该贴合面间的两个二维约束来确定基本体素方位的定位方法。该定位方式的约束集包含一个“面—面”贴合约束,记为  $C1: \text{Coincidence-FF}(F, F')$ ,其中  $F$  为参照基本体素 Ref 中的一个参照面,  $F'$  为待定方位的基本体素(称为目标基本体素) Tar 中的受约束面;以及两个二维几何约束,它们的约束元素和受约束元素分别是面

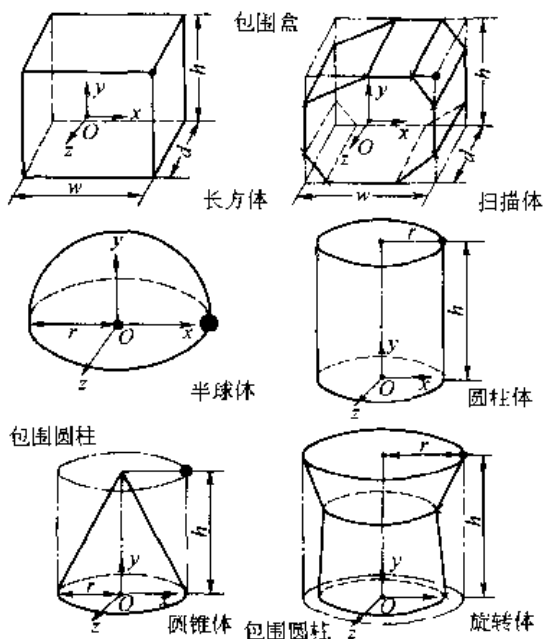


图 41.3-2 基本体素及其形状控制点

$F, F'$  的边界元素,如“边-边”距离约束、“点-边”距离约束等,用以通过面  $F$  中的约束元素将  $F'$  完全定位,从而达到通过参照基本体素对目标基本体素完全定位的目的。如图 41.3-3 所示。

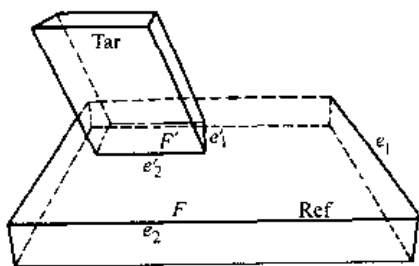


图 41.3-3 I 类定位方式

I 类定位方式

此定位方式是通过三个“面-面”距离约束(包括距离为零的“面-面”距离约束,即“面-面”贴合约束)来确定基本体素的方位的定位方法。II 类定位方式的约束集包括三个“面-面”距离约束,分别记为  $C1: Dist_{FF}(F_1, F_1'), C2: Dist_{FF}(F_2, F_2'), C3: Dist_{FF}(F_3, F_3')$ ,其中  $F_1, F_2, F_3$  为参照基本体素 Ref 或参照基本体素集(最多为三个,分别为 Ref1、Ref2、Ref3)中的三个互不平行的参照面,  $F_1', F_2', F_3'$  为目标基本体素 Tar 中的三个互不平行的受约束面。如图 41.3-4 所示。

II 类定位方式

此定位方式是通过一个“轴-轴”重合约束和一个“面-面”距离约束来确定旋转类基本体素的方位的定位方法。III 类定位方式的约束集包括一个“轴-轴”重合约束,记为  $C1: Coincidence_{ee}(1, 1')$ ,其中 1 为参

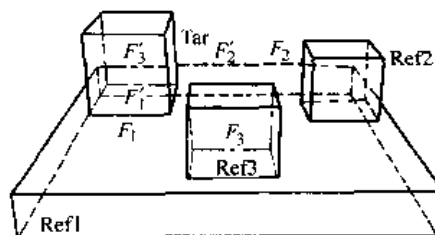


图 41.3-4 II 类定位方式

照基本体素 Ref 的旋转轴,  $1'$  为目标基本体素 Tar 的旋转轴;以及一个“面-面”距离约束(包括距离为零的“面-面”距离约束),记为  $C2: Dist_{FF}(F, F')$ 。如图 41.3-5 所示。

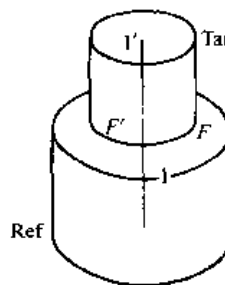


图 41.3-5 III 类定位方式

IV 类定位方式

此定位方式是通过一个“轴-轴”平行约束和一个“面-面”贴合约束和一个角度约束来确定旋转类基本体素的方位。IV 类定位方式的约束集包括一个“轴-轴”平行约束,记为  $C1: Parallel_{ee}(1, 1')$ ,其中 1 为参照基本体素 Ref 的旋转轴,  $1'$  为目标基本体素 Tar 的旋转轴;一个“面-面”贴合约束,记为  $C2: Coincidence_{FF}(F, F')$ ;以及一个角度约束,记为  $C3: Angle_{reftar}(v, v')$ ,其中  $v, v'$  分别是面  $F, F'$  的圆心,而角度则以  $v$  为原点,沿参照基本体素 Ref 的局部坐标系的  $x$  轴方向的射线与射线  $v, v'$  的夹角。如图 41.3-6 所示。

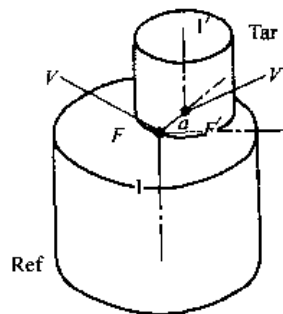


图 41.3-6 IV 类定位方式

V 类定位方式

将 I ~ IV 类定位方式不能表示的对基本体素定位的其他方法称为 V 类定位方法。

3) 定形约束 定形约束(Shape Constraints)是指

约束基本体素形状的显式几何约束。定形约束分为两类:内部定形约束和外部定形约束。内部定形约束的一个最基本特征是它的约束几何元素均为该基本体素自身的几何元素,它在基本体素实例化时自动生成。而外部定形约束指的是借助其他基本体素的几何元素来定义的作用于基本体素某一尺寸的几何约束。一般外部定形约束作为一种强制性的定形约束,只有在对基本体素完成定位后,外部定形约束才有意义,而且外部定形约束总是成对出现,共同决定基本体素的某一形状参数。

#### (4) 约束识别与求解

传统CAD系统中的几何约束需要设计者手工输入或指定,目前还没有一种几何约束自动识别的有效算法。为了提高约束识别的效率,减少设计者与VR-CAD系统的交互,必须设计一种有效的约束识别方法。下面介绍一种方法,使得几何约束采用动态输入,即用约束识别程序进行自动识别与由设计者通过交互方式进行指定共存于约束识别过程中。

约束识别主要包含两种:定位约束识别和定形约束识别。定位约束识别在基本体素的初始定位完成后启动。而定形约束识别则在设计者通过形状控制点修改已经确定为基本体素的形状时自动触发。

1) 定位约束识别 定位约束识别受到基本体素定位方式的导引。当设计者通过语音命令为基本体素制订了定位方式,完成对它的初始定位后,定位约束识别算法即自动启动。定位约束识别算法根据指定的定位方式确定所需识别的定位几何约束数目以及约束识别的效率。例如,对具有I类定位方式的基本体素而言,首先有目的地识别一个“面一面”贴合约束并有设计者对其进行确认;然后,只要在贴合的两个面之间进行有关的二维约束识别即可。

为了保证所要求的所有约束均被有效地识别出来,可以在约束识别过程中采用动态误差。以“面一面”贴合约束的识别为例,其首要条件即是在目标基本体素Tar和参照基本体素Ref之间存在平行平面。若在缺省的误差下,Tar与Ref之间不存在平行平面,此时,误差将倍增,直到约束识别程序认为Tar和Ref之间存在平行平面而成功中止,或误差过大超过某一阈值而以失败告终。

动态误差的好处是能提高约束识别的效率。但同时也带来了不一致或无效的定位。为此需要对约束进行有效性检查,这些检查应在约束求解前完成,以确保约束求解结果的正确性。关于如何设计有效性检查这里不详细说明。

2) 定形约束识别 定形约束在设计者通过形状控制点修改已精确定位的基本体素的形状时自动触

发。尽管定形约束需要一个约束对,但由于其中之一一定是定位约束中的某一个约束,因此,只需额外再识别一个约束即可。

3) 约束求解 为了提高约束求解的鲁棒性与效率,针对不同的定位、定形约束集应分别设计特定的约束求解算法。定位方式的求解算法通常采用解析方法,并在求解过程中尽可能建立有关坐标系间的相对变换以提高求解的效率与质量。关于求解过程这里不详细介绍。

### 3 VR-CAD中的多通道技术

VR技术的交互、沉浸和构想特性要求的虚拟环境给人的感受等同其在真实环境中的感受。从人的生理感知角度,人若能与环境自然交互,并且多种感官能同时受到环境信息激励时,人将有真实环境中的沉浸感。如人的眼睛、头、手的位置变化以及语音作为交互输入,相应地,人同时得到视觉、听觉、触觉和力感的反馈,这就称为多通道技术。

要使人沉浸在虚拟环境中,必须尽量把人的感官通道从物理空间隔离起来,而由虚拟环境提供感知信息。如宽视场角的大屏幕投影以及头盔显示器为眼睛提供虚拟现实环境中的图形图像;耳机隔离了物理环境中的声音而提供虚拟环境所需要的听觉信息。数据手套和头部跟踪设备可作为交互输入,抗荷服和数据手套可使人得到力反馈等。

#### 3.1 三维鼠标

拾取所使用的主要输入设备是三维鼠标,它有6个自由度,能提供鼠标的位置的方向参数。实际上三维鼠标就是一个跟踪器。为了在计算机屏幕上准确地反应三维鼠标的方位信息,进而方便设计者利用三维鼠标在三维设计空间中操纵物体和进行三维导航。三维鼠标传给计算机的方位数据在使用前必须进行一些处理。

首先,由于三维鼠标的发射装置发出的超声信号有一定的覆盖范围,当接收装置的活动范围超越其覆盖范围,此时的方位数据实际上是无效数据,必须过滤掉。

其次,三维鼠标提供的是有关方位的绝对坐标值,而对于设计者来说,更关注相对坐标,因此,需要将三维鼠标的绝对物理坐标变换到世界坐标系,映射为相对逻辑定位。

同时还要定义一些操作方法,比如鼠标复位、屏蔽操作、三维锁定等。

#### 3.2 三维物体选取机制

虚拟环境中的物体选取机制迥然不同于传统的基于二维鼠标的物体选取机制。物体选取通常需要利用

虚拟鼠标或者虚拟手,简单的方法是实时地判断这些虚拟输入设备是否与物体发生碰撞,或者输入设备的发射线、姿态以及其他的定义与物体的选取点形成交点,但是这些方法一般需要大量的运算,同时目前的求解算法并不成熟,因而求解效率不高。

可以借助于其他的设备以及在模型的算法本身上进行改进,而达到方便、灵活的选取。

一种方法是:设计者通过语音命令发出选取物体的指令,系统即进入物体选取状态,所有物体的选取点都显示出来供设计者参考,同时会有声音反馈,让设计者知道正处在物体选取状态。选取点的状态可以和视点结合起来,比如,离视点最近的选取点高亮度,移动鼠标,设计者在三维物体间穿行,当设计者按下鼠标,即确认,或者用语音命令确认。

还有一种方法是:结合语音和数据手套的选取点,即语音命令发出选取物体的指令,系统即进入物体选取状态,所有物体的选取点都显示出来供设计者参考,每个选取点都有标识,用语音选取选取点的序号(al),系统切换视图,如果选取点较近,直接利用虚拟手拾取即可。

### 3.3 三维菜单的设计

三维菜单,有时称为虚拟菜单,是目前虚拟环境中的一种较为典型的交互方式,它的存在从某个角度说是因为要适应人们以前的习惯。这点曾经成为人们争论的话题,因为与它交互不像在物理世界中那样自然。但大多数人的意见倾向于,对于抽象的命令而言,自然的隐喻(Natural Metaphors)并不一定合适。而三维菜单提供了一种与虚拟环境交互的强有力的手段。图41-3-7为一个三维菜单。

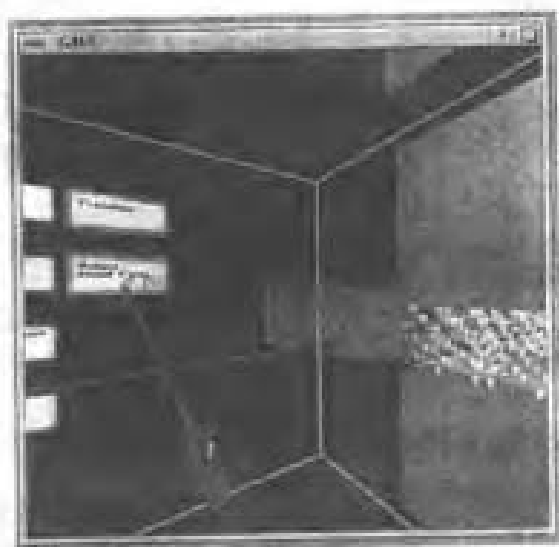


图 41-3-7 三维菜单

三维菜单是由文字或者图标组成的一个顺序表,

按照一定的几何形状排列而成。通常在虚拟现实环境中是隐藏的,当需要时,由语音命令或者其他触发激活三维菜单。在虚拟环境中与菜单的交互一般是通过数据手套和三维鼠标来完成的。

三维菜单与二维菜单的显著区别有:

1) 定位 由于在虚拟环境中,设计者可能位于三维空间的任何地方,三维菜单也是一样,这就要求三维菜单应该适当放置,才可以使用户有效地观察到三维菜单。

2) 菜单的交互 传统的二维菜单的交互方式不适合于虚拟环境中的三维菜单,靠鼠标的点击势必效率不高。菜单是三维,可以和应用环境结合起来,尽可能和应用目的直接结合起来。

3) 菜单的配置 如果允许菜单在空间中浮动,设计中可以在不需要的时候将其移到视线外去,但是当设计者需要时,可能又找不到,采用方法将菜单隐藏起来,当设计中需要的时候,及时唤醒。

### 3.4 语音系统

#### (1) 声音处理

在虚拟现实环境的声音信息中,三维空间声可为用户提供导航、状况警示(Situation Awareness),以及目标机的位置和运动方向,所以声音的定位是虚拟现实环境中声音渲染的主要内容,而现实环境中的反射、混响等环境声音特性并不明显,在此不考虑。

为使耳机提供精确的三维声音信息,涉及到三项技术:三维双耳声的建模;数字滤波器的设计;以及仿真声音运动的时变滤波器的设计和交叉衰减技术;快速卷积算法,以及长时间声信号处理。

1) 三维双耳声的数学模型 传统的立体声只能给听者一维空间感(左/右),即使多个扬声器的环绕立体声只能提供二维的声音信息(左/右,前/后),而虚拟环境需要三维空间声才能给人全方位(Omnidirectional)的听觉信息。从物理声学角度,应建立双耳三维声的数学模型。当人在真实环境中感知声音信息时,实际上处于一个由声源、传播途径、接收者共同组成的声传播系统中。从系统分析的角度,这样的系统可用图41-3-8表示。其中, $x(t)$ 是空间中某位置的声源发出的信号; $h(t)$ 是声音传播环境的脉冲响应,反映了环境的声音反射、混响特性以及环境中材料的声特性; $f_l(t)$ 和 $f_r(t)$ 分别是人的左、右耳对来自各个方位的冲激声信号的脉冲响应,即头部相关传递函数(HRTF, Head-Related Transfer Function); $y_l(t)$ 和 $y_r(t)$ 分别是人的左、右耳的耳鼓处所感受到的声音信号,人的听觉神经可根据这两个信号感知声源的空间位置。在声学领域, $y_l(t)$ 和 $y_r(t)$ 称为双耳声(Binaural

Sound)。

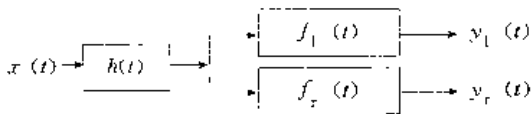


图 41.3.8 双耳声的传播、感知模型

2) 头部相关传递函数的测量 关于头部相关传递函数 HRTF 的测量可参阅相关专业书籍,此处不作详述。

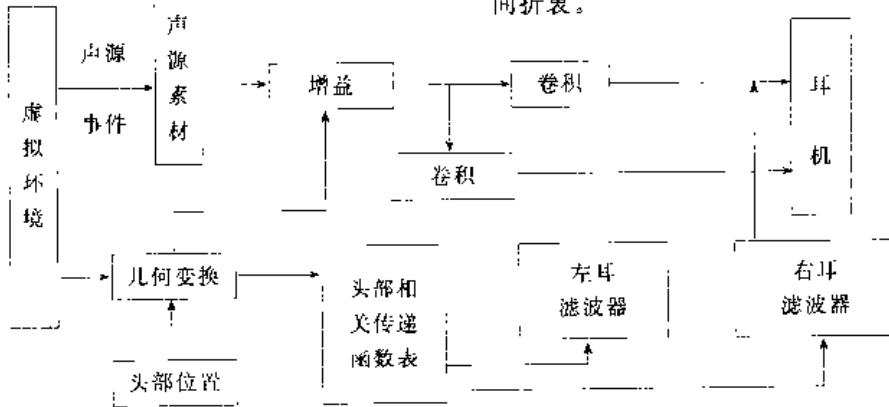


图 41.3.9 虚拟现实环境三维声音生成方案

(b) 离线的预处理工作

① 声音文件的录制、分析与存储 可从 CD、录像带、磁带上或在飞机现实环境现场录制所需要的声音文件,然后根据傅里叶分析,并结合人的听觉,对声音文件进行编辑,确定声音文件的长度,即声音信号的持续时间,把各种声音素材存在硬盘上,在系统初始化时,调入内存,以供各种声音事件的实时调度。

② HRTF 表的预处理 由于测量的 HRTF 是已知的,可以预先把测量的 HRTF 通过 FFT 算法,将其从时域变换到频域。

(c) 三维双耳声流水线实时生成

① 获取虚拟物体的状态,启动固定位置的背景声音。

② 获取环境中与声音信息相关的事件信息,从而确定所需的前景声源。

③ 获取声源位置、人的位置,以及跟踪系统采集的头部位置。通过几何变换计算声源相对人头部的方位和距离,后者控制声源的增益强度。

④ 查找该相对方位对应的 HRTF,若未测量,需要插值生成。

⑤ 在频域内,声音数据块和数字滤波器(HRTF)进行乘积处理。

⑥ 把乘积结果经逆傅里叶(IFFT)算法变换到时域,得到立体声数据块,送后台播放到耳机。

(2) 语音命令

虚拟环境中的语音命令,如语音导航、语音拾取往

3) 声音信号的卷积处理 关于声音信号的卷积处理可参阅相关专业书籍,此处也不作详述。

4) 虚拟现实环境中三维声音生成方案

(a) 一种三维声音生成方案

在虚拟现实环境的三维声音生成系统中,如图 41.3-9,采用对数据块的卷积处理来生成三维立体声。声音块的处理、生成在前台,后台通过声卡播放。需要注意声音数据块大小的选取,在前台生成、后台播放之间折衷。

往能提供简单、自然的交互手段。它使得设计者无须使用双手,将交互的手段增加了一种通道。

虚拟环境下的语音命令,其实就是一个语音识别的问题(Voice Recognition)。用户通过话筒,将声音传入系统,经过系统的语音识别功能将声音解析,大概了解用户语音的含义,然后与系统定义好的命令进行匹配,达到语音控制的功能。

在虚拟系统中,它不可能理解用户任意无关的话,因为这样会影响系统的实时性和效率。在语音识别时,系统需要解析用户的话,挑出关键词,然后按照定义好的命令,实现虚拟交互动作。一般地可以让系统设定一个专家系统,从用户的语句中推理得出其真实的想法,这样首先需要对用户的语句进行过滤、分解,通过专家系统整合成完整的一条语句。比如,用户发出:“建立一个边长为 5 的立方体”,系统将解析为“create + cube + length + 5”,通过专家系统整合为: create cube, cube length = 5。

目前已有分别在 NT 平台和 Unix 平台下的商品化语音识别软件,比如 Microsoft Speech API, IBM Viavoice。

### 3.5 触觉和力觉反馈系统

虚拟现实技术借助现代计算机技术和传感技术,让用户沉浸到计算机虚拟的空间中并与其中的人或物体产生及时互动,从而达到与虚拟环境沟通或实施控制的目的。

触觉和力觉是两种不同的感觉。触觉内容十分丰富,但是迄今触觉反馈装置仅仅能提供最基本的“触到了”的感觉,无法提供材质、纹理、温度等感觉,在虚拟设计中应用很少,且都停留在实验室阶段。

力觉反馈系统目前主要采用操纵杆时力反馈装置,通过机械臂传递给操作者力和力矩,力反馈系统一般对力经过适当缩放,主要应用在虚拟雕刻、装配时仿真。

#### 4 VR-CAD中的可视化技术

在基于虚拟现实技术的新一代CAD技术中,如何显示几何模型,进行可视化,是提高系统交互性能最重要的问题。一般地,VR-CAD系统中的可视化技术涉及到三个方面:

1) 真实感图形实时绘制 传统真实感图形绘制追求图形的真实感和高质量,对每一帧绘制速度不限制,对于VR-CAD系统,是一种必须在给定时间内,完成对场景的绘制,否则设计的沉浸性会得到巨大的削弱,但是过度的渲染又影响了设计本身关注的内容。

2) 模型的多面体表示 VR-CAD系统中,几何造型过程是模型从无到有的不断复杂动态过程。设计者不可能事先对几何模型进行渲染、着色,而是要在造型的过程中,及时地生成几何模型的多面体表示,以备渲染。如何将几何模型转化为多面体模型的算法成为影响系统显示的重要方面。对一个待绘制的几何模型而言,如果采用过多的细节去表示其外形,势必降低显示速度,影响系统的交互实时性。而过于简略的细节程度又影响了身临其境的感觉,失去了虚拟现实的内涵,所以要结合现有的软硬件条件,达到系统交互性与真实感间的平衡。

3) 大规模复杂场景的漫游 CAD场景通常由大量多边形组成,且存在大量可见多边形,任何高端图形系统也难以达到场景的实时绘制,需要对复杂场景数据进行高效管理和调度。

##### 4.1 VR-CAD中真实感图形实时绘制技术

###### (1) 消隐

真实感场景生成中,必须在给定视点和视线方向上,确定场景中哪些物体表面可见,仅对这些可见物体表面进行各种计算和变换,就可以极大减轻系统计算。所有涉及到隐藏面消除算法都涉及到景物表面距离视点远近的排序。一个物体离视点越远,它越有可能被另一个距离视点较近的物体遮挡。因此消隐可以看作一个排序问题,消隐算法效率很大程度取决于排序的效率。

消隐算法按照其实现方式可以分为景物空间消隐

算法和图像空间消隐算法两大类。景物空间消隐算法直接在景物空间(摄像机坐标系)中确定视点不可见的表面区域,并将它们表达成同原表面一致的数据结构;图像空间(屏幕坐标系)消隐算法以屏幕像素为采样单位,确定投影于每一像素的可见景物表面区域,并将其颜色作为该像素的显示光亮度。

1) 景物空间消隐算法 在景物空间实现的消隐算法主要有:表优先级算法,BSP算法,Weiler-Atherton算法。

2) 图像空间消隐算法 在图形空间实现的消隐算法主要有:Z缓冲器算法,扫描线算法,光线追踪算法。

消隐涉及到对场景中景物的排序,是最为耗时的部分。需要考虑利用相邻帧画面的可见性、连贯性及图像、景物空间连贯性等实时消隐算法来加速消隐的过程。目前主要方法有:

① 层次Z缓存算法 将明显不可见面片和需计算面片用包围盒的方法层次化,通过建立的一棵消隐树,共享许多消隐计算和判定的结果。主要创新在于引入屏幕包围盒树来快速判别景物中明显不可见面片。

② 可见性预计算方法 采用预处理技术,将一些计算量大的信息预先计算好并存储起来,绘制时只需要调用这些信息即可实现对场景的快速绘制。在预处理时,首先将景物空间剖分成一系列网格,然后对每一网格,计算当视点位于该网格内时沿每一视线方向可能可见的面片集合(PVS),并存储在数据结构中,当视点在该网格内移动时,场景的绘制可以简化为对该网格中记录的沿给定视线方向潜在可见面片的绘制。

###### (2) 光亮度

光亮度计算是真实感图形生成的重要环节,它决定了画面最终视觉效果。光亮度计算和光照明模型是密不可分的。光照明模型主要有Lambert模型、Phong光照模型等。

图形学中常用增量方法来计算光亮度,主要有两种方法:

1) Gouraud明暗处理 Gouraud明暗处理将曲面表面的光亮度取为近似表示该曲面的各多边形的顶点光亮度的双线性插值。其保证由多边形近似表示的曲面上各处光亮度的连续变化,但在相邻多边形公共边界上光亮度的二阶导数不连续,产生马赫带效应。

2) Phong明暗处理 也称为法向量插值明暗处理,能够正确模拟高光,可以大大减轻马赫带效应。其对多边形顶点处法向量做双线性插值,在多边形内构造一个连续变化的法向量函数,将依据此函数计算的多边形内各采样点的法向量带入光亮度计算公式,得到多边形近似表示的曲面在各采样点处的光亮度。



## 4.2 VR-CAD 中多细节程度模型生成技术

### 4.2.1 网格简化

对于虚拟现实系统,维持一个稳定的具有交互式特点的帧速率是最重要的。虚拟场景依赖稳定的帧速度使人的活动融入到虚拟环境中,以保证人机交互。不稳定的帧速率,特别是帧与帧渲染之间延迟太长,会破坏可视化的连续性。然而数据量本身非常庞大,对其进行可视化本身需要大量的 CPU 时间和内存,同时对虚拟场景的操作也增加了系统的负担。因此如何提高实时性成为可视化面临的难题。

物体的几何模型常用多边形网格表示(称多面体模型),最典型和最常用的就是用三角面片拟合的多面体模型。一个自然的解决办法就是在不致影响视觉效果或保证物体视觉特征还可以接受的前提下,对物体多面体网格模型进行简化处理,使用较少的三角面片来表示物体。国内外学者已提出了多种网格简化算法,其中以 Hoppe 的累进精简算法最为广泛使用,但是这种精简方法大多用于对静态几何模型进行精简。网格精简算法的类型主要有以下几类:

#### (1) 拓扑结构保持型算法/拓扑结构非保持算法

1) 拓扑结构保持型 必须保持模型的拓扑结构(包括局部拓扑结构以及全局拓扑结构),模型的精简不会精简到很低程度,大多数算法对输入的模型有要求,如果模型包括退化情形(如多边形重叠、边由 2 个以上面片共享等),则不能进行处理。绝大多数的简化算法都属于此类。

2) 拓扑结构非保持型 不保持模型的局部或全局拓扑结构,精简模型只要其图形生成效果和原模型一致,则可以不保持原模型的拓扑结构。这类算法约束条件较少,因而简化程度较高,在要保持较高帧速率且视觉效果可以一定程度下降时,可以使用这类算法,但是这种精简可以将一个复杂的城堡简化成 1 个面片。

#### (2) 自适应细分型/采样型/几何元素删除型

1) 自适应细分型 这种方法首先建立原始模型的最简化形式,然后根据一定规则通过细分把细节信息增加到简化模型中,从而得到较细的 LOD 表示。这种方法不常用,因为在一般情况下构造最初网格的最简模型相当困难,主要用于均匀的网格,如高度场。

2) 采样型 这种方法类似图像处理中的滤波方法,有时不能保持拓扑结构不变。它对原始模型的几何表示进行采样。一种方法是从它的表面选择一组点,另外一种方法是把一个三维网格覆盖到模型上并对每个三维网格单元进行采样。在另一种方法的选择一组点时,又可分为随机点或根据一定规则选择特征点,显然

选择特征点的方法的近似程度高,但计算量大。

3) 几何元素删除型 通过重复地把几何元素(点、边、面)从三角形中删除来得到简化模型。这种删除一般有三种情形:一是直接删除,一是通过合并两个或多个面来删除边或面,还有一种是对边或三角形进行折叠。删除操作一直进行,直到模型不能被简化或达到用户指定的近似误差为止。绝大多数算法要求对几何元素进行删除,需要判断几何元素是否可以删减,不要破坏模型拓扑结构。

#### (3) 局部算法/全局算法

全局算法对整个模型或场景模型的简化过程进行优化,而不仅仅根据局部的特征来确定删除不重要的几何元素。主要方法有 Turk 的表面上重新分布顶点的方法;Hoppe 使用全局能量最小化方法,Cohen 提出包络网格以及 Eck 的小波表示方法。

局部算法应用一组局部规则,仅考虑模型的某个局部区域特征对模型精力学简化,主要方法有 Schroeder 的顶点删除, Rossignac 的顶点折叠, Hoppe 和 Cuezic 的边折叠以及 Hamann 的三角形删除。

1) Schroeder 的顶点删除法 基本思路是指定一个最小的距离阈值,如果模型中某顶点到由该顶点定义的平均平面的距离小于该阈值,则删除该顶点,并采用递归循环分割法对删除顶点后遗留的空洞进行三角剖分,通过调整距离阈值大小可生成层次化模型。Schroeder 将该技术应用于从医学 CT 数据中抽取的等值面模型及地形模型的简化,大量消减了原模型中的三角形数,同时保留了原模型的几何特征。Schroeder 算法简单,执行效率高。

2) Turk 的重新布点法 基本思路是指定一个新模型所包含的顶点数,首先将这些点布置在曲面上,原则是面积大的多边形内多布一些点,曲率变化大的多边形内多布一些点,新点集合中可以包含原模型中的点;第二步生成由新旧顶点共存的网格,即将新点插入到原模型中,修改原模型网格;最后删除模型中不在新点集中的顶点,得到由新布点集合中的顶点组成的简化模型。通过调整新模型中的顶点数,可以生成层次化模型。这种方法仅适用于光滑曲面,且简化模型中引入了新点。

3) Hoppe 的能量函数法 其中能量函数由三部分组成:距离能量、表示能量及弹簧能量。其中距离能量反映原始顶点集与简化模型的距离偏差。该能量越小,表明简化模型对原始模型的逼近精度越高。表示能量定义为表示因子  $Crep$  与模型顶点数  $m$  的乘积, $Crep$  值越大,表明模型表示的简洁性越重要, $Crep$  值越小,表明对原模型的逼近精度要求越高,因此通过指定不同的  $Crep$  值,可以控制模型的复杂度,构造层次

化模型。这种方法的特点是用能量函数的变化指导网格简化,通过在能量函数中加入一项表示能量将网格简化视作一个网格优化过程,通过能量函数中的距离能量变化反映出简化后的模型对原始模型的逼近程度。Hoppe 给出了对三维扫描仪测量的数据模型进行简化的实例,效果十分理想,但算法的执行效率很低。

4) Hinker 的合并共面多边形法 通过找出最大法矢夹角在某一给定值之间的一组多边形,将其看作近似共面的多边形,把这组多边形合并成一个多边形,对合并后的多边形进行三角剖分。

5) 顶点簇(Vertex Clustering)方法、Hamann 的三角形删除法、Rofard 的边退化法、基于八叉树表示的模型简化方法、基于简化信封(Simplification Envelope)的模型简化方法、基于感知系统的模型简化方法、基于超曲面(Surface)的模型简化方法、基于体素表示的模型简化方法等。

6) Renzen 提出了一种通用的非结构化网格简化方法 特别是解决了四面体网格的简化问题,即体简化(Volume Decimation)方法。体简化比面简化(Surface Decimation)难度要大,因为面简化过程中,删除一个顶点,与该顶点相连的顶点可以按逆时针排序;而体简化过程中,删除一个顶点,包围该顶点的若干三角形面片无法进行排序,因此一般的面简化方法无法直接推广到体简化。Renzen 的方法实际上可分为两步,第一步,即对删除顶点后遗留的空壳体进行四面体剖分;第二步即解决剖分后存在的拓扑不相容问题。

#### (4) 其他方法

还有一些其他分类方法,如视点相关性、误差可控性、实时性和连续过渡性等特征。

1) 视点相关性 把算法分为两大类,即与视点无关的模型简化算法和与视点有关的模型简化方法。早期的算法都与视点无关,近两年出现了一些与视点相关的方法,这是一个重要的发展趋势。

2) 误差可控性 有两层含义,一是指用户对整个模型的近似误差是否可以控制(全局);二是指用户对局部误差是否可以控制。进一步讲,用户可以有选择地对模型的不同部分使用不同的误差度量。

3) 实时性 模型简化的目的就是为了加快绘制速度,达到实时图形生成。这种分类方法有一定的模糊性,原因是实时性与所使用的计算机的运算速度有关。

### 4.2.2 多分辨率模型生成

高精度的扫描测绘手段为复杂物体基于多边形网格表示的三维几何建模提供了新的高效手段,但由于采样精度高,由此建立起的三维模型的复杂程度远远

超过了当前计算机实时的图形处理能力。如何降低这些模型的复杂度,减少图形系统需处理的多边形数目,实现实时交互,已经成为计算机图形学研究中的一个重要课题。为此人们提出了各种方法,细节层次 LOD (Level of Detail)便是其中一种非常有效的控制场景复杂度的方法。

所谓 LOD 技术,就是在实时显示系统中所采用的细节省略(Detail Elision)技术。这项技术首先由 Clark 于 1976 年提出,基本思想是:如果用具有多层次结构的物体集合描述一个场景,即场景中的物体具有多个模型,其模型间的区别在于细节的描述程度,那么实时显示时,细节较简单的物体模型就可以用来提高显示速度。实时显示时,模型的选择取决于物体的重要程度,而物体的重要程度由物体在图像空间所占面积等多种因素确定。在计算机图形学中,场景中的物体通常是用多边形网格描述的,因此 LOD 模型的自动生成就转化为三维多边形网格的简化问题。

LOD 模型的缺点是所需存储量大。当使用 LOD 模型进行绘制时,有时需要在不同的 LOD 模型间进行切换,这样就需要生成多个 LOD 模型。此外,离散 LOD 模型无法支持模型间的连续过渡。为此,人们开始研究多分辨率模型。

严格地讲,多分辨率模型是指一种紧凑的模型表示方法,从这个表示中可以生成任意多个不同分辨率的模型,一个典型的代表是 Microsoft 公司的 Hoppe 提出的累进网格。不过,由于有些网格简化方法能够生成连续的 LOD 模型,因而在一些文献中,也把这类模型统称为多分辨率造型。LOD 的关键技术不外乎以下问题:

1) 数据的存储布局 数据在内存中的布局必须要方便算法的实现,同时最好还要降低操作系统缺页中断的次数,也就是降低内外存之间的数据交换的次数。

2) 如何在生成连续的 LOD 化的模型网格 在模型 LOD 化过程中,要让两个由不同层度的细节的区域之间能平滑的过渡。

3) 节点评价系统 这个系统必须要使生成的网格能尽量的减少几何形变,尽量使画面质量能接近全分辨率时候的模型,同时还要保证实时性。

### 4.3 VR—CAD 系统中的复杂场景实时漫游技术

三维导航的目的是使设计者能方便地改变观察产品模型的方位,以对产品模型的某一局部进行细致的观察、分析,可以深入到模型的内部,查看模型的内部细节。一个恰当的导航机制有利于设计者快速、自然地达到想去的任何位置。

### (1) 相机模型

相机模型在计算机图形学上已有相当深入的描述,但虚拟现实技术的相机模型,要考虑到两眼的视图。

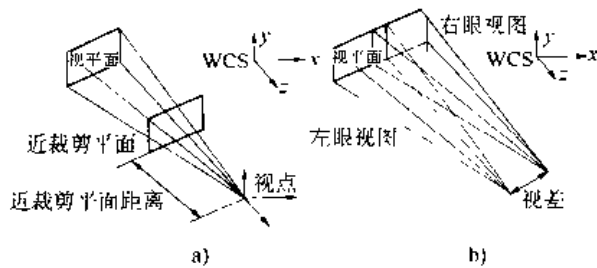


图 41.3-10 相机模型

a) 单眼视图 b) 双眼立体视图

图 41.3-10 描述的为相机模型图,其中除了单眼视图的所有参数外,增加了左右眼的视差(Parallax)参数,对于非对称立体视图,还要有收敛距离(Convergence Distance)等参数。

其中相机模型的主要操作和语义见表 41.3-1。

表 41.3-1 相机模型的主要操作和语义

相机操作	语 义
Dolly: 移动摄影机	目标点不变,沿视线方向移动视点
Orbit: 盘旋	目标点不变,视线绕目标点旋转
Track: 来回移动镜头	同时移动视点和目标点
Roll: 翻滚	视点和目标点不变,改变转角
Zoom: 推拉镜头	视点和目标点不变,转角不变,只改变视角,分为推镜头(Zoom In)和拉镜头(Zoom Out)

### (2) 三维导航机制

设计并建立了高效的场景数据树后,为便于设计者在虚拟环境中方便地调整观察方位,防止在虚拟环境中迷失,可以采用基于语音、三维鼠标、虚拟手套等设备的三维导航操作:

1) 设置缺省相机参数(Default) 在设计某一具体的模型时,总有一个观察位置和方向较适合于该模型的观察和分析,可以将此状态下的各种参数,设置为相机的缺省参数。这种参数的获取一般可以采用计算或者在虚拟环境中的漫游结果取得,用语音输入到系统中。

2) 相机缺省参数的重置(Reset) 缺省的相机参数可以非常方便地观察、分析物体,但用户在设计中需要经常转换位置和视角来建模,当视角变化后,用户可能迷失在虚拟环境中,需要重新回到初始状态,这种功能非常重要。

3) 设置地标(Landmark) 如果设计者觉得某一观察位置、方向较有利于对模型的观察分析,可以将其设置为地标。地标的作用是储存当前的视点、视线方向等相机参数,使设计者很容易获得某一位置的视角模型。

4) 基于三维鼠标的漫游 三维漫游对于 VR-CAD 来说,非常重要,它能让用户自由地在虚拟环境

中移动,从不同的视角观察设计的模型。三维漫游的本质其实就是在三维空间如何操纵立体相机模型。利用三维鼠标多自由度的优点,可以这样来实现三维空间的漫游:语音命令发出三维漫游指令,就可以通过移动三维鼠标控制视点。定义好三维鼠标的各个键的功能,获取用户按的键,确定用户漫游的方向,漫游速度可以用语音命令来给出。

5) 基于语音的漫游 设定语音操作语句,解析后即可用于漫游。比如发出开始漫游命令,可以定义前进、后退、左转、右转,还可以精确到前进多大距离,转动多大角度,就可以方便地在三维空间中漫游。但是这种方法,语音操作语句较多,实时性将受到考验。

6) 基于手势的漫游 结合语音操作来确定漫游命令和漫游速度等,用手势来确定漫游方向、视角等,因为手势的命令需要识别,系统的实时性也要求很高。但由于设计者能够依其自身的意图实时地在三维虚拟环境中借助语音和直接自然的三维操作控制相机的运动过程,从而有利于做到真正意义的三维交互式漫游。

7) 推拉镜头(Zoom) 镜头的推拉是通过改变视角来实现的。我们将三维鼠标的往前移和往后拉对应于推镜头和拉镜头,即三维鼠标前移用于缩小视角,而后拉用于放大视角。

8) 物体聚焦(Focus) 物体聚焦功能将视线聚焦到某一选定的物体上,也即是使物体显示在视角中心。这不仅有利于对该物体进行细致的观察,而且将显著提高该物体相关的三维造型操作的效率。物体聚焦是通过改变视线方向和视角来实现的,依据物体的大小、位置和相机当前的位置,自动调整视线方向和视角的大小以使物体显示在中心。

9) 翻滚(Roll) 与 Zoom 操作类似,将三维鼠标绕某一轴的旋转角度用于控制镜头的视角,从而使得翻滚操作更加直接、自然。

10) 动态观察 通过对预先设置的地标进行插值,可以生成相机的运动轨迹,对相机的运动过程进行仿真。在动态观察过程中,设计者沿着定义的路径观察系统的物体,可以发现模型的合理性和设计缺陷。

## 5 设计中的人员因素分析

虚拟现实技术提供了一种全新的人机交互方式,在虚拟现实系统中真正占主导地位的还是人,任何一个虚拟现实系统如果忽略了人的因素,那么它只能是失败的系统,虚拟现实系统就是要缩短人机间的隔阂,试图开发出最适合于人使用的各种交互设备。

在设计领域中,人员因素分析(Human Factor Analysis)处理的即是一般的可达性问题,如工作空间(Work Space)、工作间距(Work Clearance)和零件的

可达性等。虚拟现实系统具有更好的交互能力和更逼真的临场感,更适合设计者的观察分析。

人体运动的模拟一直是计算机动画领域的研究热点和难点之一,主要因为人体的造型非常复杂,而且对其赋予真人一样的行为更是困难。在交互的虚拟现实系统中,虚拟人体模型可以用在对虚拟三维世界进行评估。

宾夕法尼亚大学计算机图形研究中心开发的JACK是目前最为著名的人体模型之一,JACK是一个用于操作有关节的人体模型的通用交互环境。其最普通的应用是,首先从CAD系统中将要分析的几何模型读入JACK中,然后引入带有关节的人体模型,并将其用于对几何模型进行配合、可达性、运动、力学、视野分析。JACK中的人体数据库采样来自于美国陆军在1988年进行的人体调查,JACK中人体模型有70个关节,共有136个自由度,Jack中通用约束引擎利用了反运动学来求解计算关节的位置。JACK最优越的地方是可以实时地检测虚拟人体模型物理行为的可能性及能力。比如,当用户指定JACK系统中的某一个人体举起一个重物,系统会计算该人体模型的状态,是否处于临界状态或者其力不能及,并且会计算人体模型的后援反应。在虚拟环境中,JACK经常用于诸如伸展空间(Reach Space)、视域(Field of View)、关节转矩负载(Joint Torque Load)、碰撞检测和碰撞避免(Avoidance)等等。JACK模型曾经用在汽车中作为人机工程分析,目前EDS公司开发了完整成熟的Jack软件,用来处理虚拟样机中的人机工程问题。图41-3-11为车间装配仿真。



图 41-3-11

波音公司的Monkey也是一个非常有名的人体模型,如图41-3-12,有18in高,6in宽。在其主要关节处放置传感器,传感器的数量取决于关节如何运动。每个传感器都有一个相关的旋转限度,每个传感器都足够宽松,不需过分用力即可平滑地旋转、移动。同时,又具有足够的摩擦力,以保证不需要其他的支撑物时

Monkey可以保持固定的姿势,而且还可以和其他设备相连。Monkey已经集成到波音公司已有的人体模型中去,用来对飞机设计人员因素分析。分析内容主要包括器具的可见性(Instrument Visibility)分析,零件的可达性分析以及在数字原型上进行的维修过程的有效性分析等。在这之前,他们进行一连串的可用性研究,以定量地对利用Monkey取代已有的基于鼠标的系统的好处进行定量分析。结果表明,Monkey及其相似设备在简化需要指定人体姿势的任务,提高生产能力上显示了相当的潜力。



图 41-3-12

IBM公司在DMU环境下创建和模拟精确的标准数字化人体模型,在产品全生命周期的早期阶段进行关键性的人机互动分析,其人体模型构造器产品2(HBR)可在DMU环境下创建和模拟精确的标准数字化人体模型,从而在产品全生命周期的早期阶段实现人机互动分析。在人体模型构造器产品2内所提供的工具包括:创建人体模型、性别与身材比例规范、人体模型操作技术、产生动画以及优秀的视觉模拟。人体模型构造器产品2(HBR)可以满足各类用户各自所具有的特定制需求,包括来自航空企业、汽车制造业、工厂设计、重型工业、船舶制造以及电子产品的设计人员、管理人员、维修人员和概念设计人员。

人体模型构造器产品2(HBR)可与人体测量编辑器产品、人体姿态分析产品以及人体行为分析产品共同高效地使用,从而实现高级的详细的人体模型创建与分析,以提升人体舒适度、可操作性以及安全性。将上述产品整合在一起便是完整的人机工程设计与审查解决方案。

德国奔驰公司的人与汽车研究小组研究了汽车内部的空间感知问题,他们利用虚拟环境试图确定颜色、纹理和几何等因素中,哪一个对人的视觉影响较大。系统采用立体眼镜/头盔、头部跟踪设备和数据手套/三维鼠标,用户可交互指定汽车的颜色或大小。

福特公司也对虚拟环境在汽车装配中的使用进行了评价,目的是评估各种装配动作是否符合人机工程原理。为此,他们将 CAD 的模型导入虚拟环境中,用户佩戴上头盔获得立体视觉,用数据手套操纵虚拟零件并将它们装配到虚拟汽车上去。系统可以检查零件和汽车间的干涉和碰撞,通过佩戴运动跟踪设备,可以对零件的可达性进行评估。

爱荷华大学的 Vujosevic 等人设计了一个虚拟拆

卸系统,它不仅可用于确定拆卸顺序,而且还可以从不同方面,包括时间、成本和人员因素,对拆卸过程进行可行性分析和评价。利用虚拟人体模型来模拟在拆卸过程中人与设备的交互过程,能对拆卸中最重要的伸手和抓取等动作进行快速分析。

可以看出:在设计中的人员因素分析将越来越重要,事实上,如飞机座舱、汽车内饰、引擎、航天设备等的设计中,人机工程分析已经占有相当重要的地位。

## 第4章 虚拟现实开发工具集

虚拟现实开发的工具非常多,本章主要介绍一种是虚拟现实的建模语言 VRML 以及目前最为流行的实时视景仿真开发工具 SGI OpenGL Performer。其他开发工具集,比如 OpenGL, Multigen Creator, VEGA, WTK 等请参考相应书籍。

### 1 虚拟现实建模语言 (VRML)

#### 1.1 VRML 的概述和发展历程

VRML (Virtual Reality Modeling Language) 是随着全球互联网的发展应运而生的。

Web 是以 HTML 为核心的网络技术,HTML 以一种标准格式将全球信息资源统一起来,并通过复杂的链接关系,构成了庞大的、内容丰富的多媒体信息空间,用户可以方便地在二维网页上进行文本和图像浏览。但是,随着网络技术的进步,人们已经不满足于二维的被动式的网页,而是需要能以自然的方式与网络中的三维环境进行交互;即将 web 从以页面为中心的模式转变为近人化的三维、交互、逼真的虚拟世界,VRML 正是在建立这种三维超媒体 web 中诞生的。

VRML 是一种描述交互式三维多媒体的标准文件格式,它被设计成为与互联网上的三维环境可集成的开放系统,并于 1997 年 12 月成为国际标准 (ISO/IEC 14772 1:1997),亦称为 VRML97。实质上,VRML 是一种三维交换格式,它定义了当今三维应用中绝大多数常见概念,如变换层级、光源、动画、材质属性、雾和纹理映射等,其基本目标是确保能成为一种有效的三维文件交换格式。它所提供的技术能够把三维、二维、文本和多媒体集成为统一的整体,当把这些媒体类型和脚本描述语言结合在一起时,可以产生一种全新的交互式应用,使三维环境模型取代 2D 桌面模型而成为基本用户界面模式成为现实。正如国际标准化组织所描述“虚拟现实建模语言 (VRML) 是用来描述三维交互物体和场景的文件格式。VRML 的设计主要用于 internet、intranet 和以太网。VRML 的目的是成为一种集成三维图形和多媒体的通用交互格式。VRML 用于许多应用领域,如工程和科学可视化,多媒体显示、娱乐和教育,网页及共享虚拟场景”。

VRML 的发展可以追溯到 1994 年。1994 年,Mark Pesce 和 Tony Parisi 创建了被称为 Labyrinth (迷宫) 的浏览器,这是 WWW 上 3D 浏览器的早期原

型。同年,Mark 和 Brian Behlendorf 创立了 VRML 邮递表 WWW VRML,并发出为 WWW 上的 3D 应用建立正式规范的倡议。SGI 的 Gavin Bell 注意到了 Open Inventor 很适合作为这样的规范,并很快组织了一份提案。经过一场激烈的辩论之后,Inventor 提案被选定为未来规范的工作文档。随后,Gavin 在很多 Inventor 工程师的帮助下,修订并最终制订了 VRML1.0 规范的草案。此规范 1994 年 10 月在第二届 WWW 国际会议上公布。之后,VRML 2.0 规范的第一版于 1996 年 8 月在 SIGGRAPH'96 会议上与人们见面。

1997 年 12 月,VRML 作为国际标准正式发布,并于 1998 年 1 月获得 ISO 批准 (国际标准号 ISO/IEC 14772 1:1997),通常称为 VRML 97。它是 VRML2.0 经编辑性修订和少量功能性调整后的结果。作为 ISO/IEC 国际标准,VRML 的稳定性得到了保证,它将推动 Internet 上交互式三维应用的迅速扩展。

VRML 协会是一个非盈利性的开放式贸易协会,其作用是定义、推动和发展 VRML 标准。协会于 1996 年创建,成员包括 3DLabs、Apple、Boeing、CosmoSoftware、Fujitsu、IBM、Intel、Intervista、Microsoft、Mitsubishi、Oracle、Platinum、Philips、Sony、Superscape 等 35 家 Internet 厂商。VRML 协会支持的 17 个技术工作组是推动 VRML 标准发展的主要技术力量。协会目前的工作重点是 VRML 的市场应用,如企业计算和数据库应用等。

目前,VRML 结构组织成立了 17 个工作组 (Working group),每个工作组负责某个和 VRML 有关的专题技术的研究和实现工作。它们的研究目标基本涵盖了主要发展动向。如人性动画工作组 (Humanoid Animation WG) 研究利用 VRML 表现人类行为特征;面向对象扩展工作组 (Object-oriented Extensions WG) 探讨对 VRML 进行面向对象扩展的方法;界面组件工作组 (Widgets WG) 为开发者和用户提供一套基础性的、可自由使用的标准用户界面组件集,并提供支持基本组件集和所有 VRML 组件的理论框架。它为网络三维运用提供了一个业界标准,并与 MPEG 等国际化标准保持着密切的联系。它的进一步完善和发展必将为虚拟现实技术软件实现的规范化奠定基础。

VRML 目前是 VRML97 (2.0 版本),具有更多扩展功能的 VRML 下一代标准是 X3D,其全面支持

XML 和 Java 技术。但是 X3D 目前还没有完全成熟, 解析它的工具也正在完善, VRML 是各软件商发布 3D 内容采用最普遍性格式, 本章节主要介绍 VRML97 (VRML2.0) 规范。

## 1.2 VRML2.0 的语言规范

在 VRML 中, 虚拟场景用场景图 (Scene Graph) 描述, 场景图的基本单元称为节点 (nodes), 节点间以“父子”关系形成层次性结构。节点之间可以通过事件相互通讯, 时间通过路由在场景图中传播。传感器 (sensor) 节点能够检测用户动作或时间推移, 从而产生初始事件, 构成交互性和动态行为的基础。利用脚本 (script) 节点可以自定义行为。内插器 (interpolator) 节点实际上是内置脚本, 提供基本动态行为的计算。内联 (inline) 节点提供了场景之间的链接能力。原型 (prototype) 类似于程序设计中的函数或对象的封装机制, 是一种高于节点的层次结构, 为构建虚拟场景提供了便利。

### 1.2.1 VRML 文件

VRML 文件是虚拟场景的文本描述, 其扩展名为 .wrl 或 .wnl, 压缩类型的文件格式是 \*.wrz。一个 VRML 文件包括四个主要成分:

- 1) VRML 文件头。
- 2) 原型。
- 3) 造型和脚本。
- 4) 路由。

其中 VRML 文件头是每一个 VRML 文件所必须的, VRML2.0 的文件总是以“#VRML V2.0 utf8”开头。VRML 文件还可以包括下列条目:

- 1) 注释。
- 2) 节点域和域值。
- 3) 定义的节点名。
- 4) 使用的节点名。

VRML2.0 场景是由一套按等级排列的场景实体组成的。这个模式类似于用于 VRML1.0 中的模式。等级是建立在父子系统关系之上的, 父母可能有许多子, 其中的一些子可能本身就是父母。

程序 1 是一个完全的 VRML2.0 的例子, 这个例子很简单, 并不需要太多的背景知识, 但是要你对 VRML2.0 场景中的基本概念有一个足够的了解就复杂很多。

```
程序例 1 一个简单的 VRML2.0(chair.wrl)
VRML V2.0 utf8
#
# chair.wrl(chair)
```

```
#
DEF CHAIR Transform {
translation 0 2 10
children [
Transform {
translation 0.5 0.5 0.5
children [
DEF SEAT Shape{
appearance Appearance {
material Material {diffuseColor 0 10}
}
geometry Box {size 1 0.1 1}
}
]
}

DEF LEG Transform {
children [
shape {
appearance Appearance {
material Material {diffuseColor 0 01}
}
geometry Cylinder {height 1 radius 0.1}
}
]
}

DEF LEG2 Transform {
translation 0 0 1
children[
USE LEG
]
}

DEF LEG3 Transform {
translation 1 0 0
children [
USE LEG
]
}

DEF LEG4 Transform {
translation 1 0 1
children [
USE LEG
]
]# end of children for chair
]# end of transform
```

```

DEF BACK Transform 1
translation 0.5 -1 0.7
rotation 1 0 0 1.2
children [
USE SEAT
]
}

```

例子显示,一个简单的椅子由椅面,四条腿组成。这个椅子也许不会是世界上舒适的椅子,但它将为我们而工作。注意这种方法并不是建立在这种模式的最佳方法。之所以建立这样一个椅子是为了强调 VRML2.0 场景中的某些方面的问题。

图 4.4-1 给出了这个模型在 CosmoPlayer 2.0 的浏览器的模样。



图 4.4-1

### (1) 结点

VRML2.0 的基本元素是结点,结点是组成 3D 场景的基本元素。每个结点被类型化并包含有一系列使结点参数化的字段。反过来,每一种字段是一种类型,VRML2.0 明确定义了类型和字段的默认值。

VRML2.0 包含有 50 个结点,大约比含有 1.0 中的多 20 个。这一节和下一章将利用一些新的结点。大部分结点被用于使场景作者能够把一些动作加到它们的场景中去。

在上一节的例子中(程序例 1),几何上 Box 结点表示椅子和椅背,用 Cylinder 结点表示椅腿(15 行和 26 行)。注意,不管怎样,这些几何结点包含在 Shape (形状)结点中。反过来,它们又包含在 Transform 结点中。这种“包含”就是我们定义等级的方法。

1) 命名结点 结点是以 DEF 为前缀命名的,它分配一个字符串给一个特定的结点。把 LEG 分配给 Transform 结点,这其中包含了椅子的几何形状。例如,你将会用程序例 2 中的代码表示椅子形状。

程序例 2 Transform 结点 DEF LEG Transform

```

{
...
}
}

```

DEF 构成了一个分享模式的基础,通过它,许多命名过的结点都可以在以后的某些场景中通过 USE 语法调用。仔细看程序例 1,你就会发现椅腿被重新创建,椅腿被定义(DEF)一次,并且被其他腿调用过三次。

2) 组结点和子结点 结点有两种类型,组结点(Grouping)和叶结点(leaf)。组结点,就像 Transform,组成子结点,子结点有可能是叶结点或别的组结点,组结点为子结点提供子结点的坐标空间,这个坐标空间和父结点的坐标空间有关。这将使得场景作者能够建立结点等级,这种等级是随着场景等级下降累积的结果。

再看程序 1,你可以看到定义了两个等级。第一个等级……用 Transform 结点定义且命名 CHAIR……从第五行开始且有五个子,每一个子都有一个 Transform 结点。第一个子结点是对椅面的描述,第二个子结点是对一个腿的描述,剩下的三个腿被用来描述其他三个腿。

第二个等级(第 50 行)很简单的,它建立在 Transform 结点上且只包含一个子结点,这个子结点描述了椅背,它是重复用 SEAT 结点来创建椅背。

第一个等级表明父与子之间的关系。最初的名 CHAIR 的 Transform 结点确定了一个沿 X 轴 2m 和 Z 轴 10m 的坐标转移,因为这个结点的子从它们的父处继承了坐标空间,因此它们以这个坐标作为它们新的起点,子定义的任何转移都与父的坐标空间有关。

除了 Transform 结点,还有 4 个组结点,Anchor, Billboard, Collision 和 Group。这些结点和 Transform 结点扮演了相同的角色,并且为它们的子提供了等级转移的空间。

3) 叶结点 叶结点像 Shape 一样把一个实体放入一个坐标空间中,这个坐标空间是由周围组结点定义的。Shape 是一种新的 VRML2.0 结点,它只包含场景中可见部分的几何和外形信息。Shape 结点被设计用以支持场景中简单的共享问题,这样使得场景组成 VRML1.0 简单。再回到例子中,一个简单的形状,像椅子这样,可以被定义并且可通过 DEF/USE 在场景中调用。

几何字段实质上是一个结点指针,它包含的结点描述了物体的几何特征,例如,一个锥体结点或一个球体结点,表象字段使得作者能定义如颜色和亮度这些可视的几何特征。

### (2) 场景



VRML 文件定义组成等级的一套结点。场景就是这样 一个名字:它被用来解释这些出现在 VRML 文件中的完全有顺序的场景等级。每一个都是一个单独

的有根结点,同时有一系列子结点。

一个场景图如图 41.4-2a 所示,更具体的讲,图 41.4 2b 指出了椅子这个例子的场景等级。

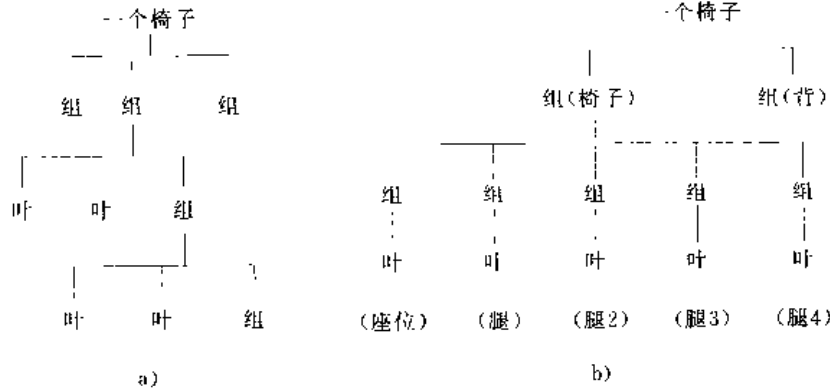


图 41.4-2

### 1.2.2 VRML 的节点及用法

VRML 中的节点是其最基本的成分,不同节点的组合构成了虚拟世界的场景图。VRML2.0 共定义了 54 种节点类型,它们可分为以下 7 组:

- 1) 形状节点:定义几何形体。
- 2) 属性节点:定义图形绘制方法。
- 3) 编组节点:将若干个节点组合成一个节点。
- 4) 传感节点:检测用户的行为,提供交互能力和动态行为。
- 5) 内插节点:为虚拟世界提供关键帧动画功能。
- 6) 脚本节点:使编程人员可以编写自定义的行为。
- 7) 其他节点:提供背景、雾化、光照、声音等环境效果,控制场景细节、用户视点以及提供空间细节等。

利用 VRML 的众多节点,可以设计交互的三维物体和虚拟场景,并通过纹理、光照等处理手段使虚拟世界更具有真实感。

#### (1) 形状节点及造型的平滑绘制

VRML 提供几种原始几何节点,将它们作为 Shape 节点的域值能够创建原始造型。这些原始造型包括 Box、Cone、Cylinder 和 Sphere 节点,每个原始几何节点具有一个或多个域,便于设定其几何形状参数。在缺省状态下,这些原始造型均是实心体,但 Cone 和 Cylinder 节点中含有是否描绘底面和侧面的控制域,通过改变赋予这些域的布尔逻辑值可以创建空心造型。组成圆锥体、圆柱体和圆球体曲面的面数由 VRML 浏览器决定。一般情况下,一个圆柱体的侧面由 18 个或更多的面组成,一个球体有 60 个或更多的面组成。为了保证实时绘制虚拟环境,浏览器通过计算观察者与造型之间的距离,对原始造型的面数进行优化控制:当观察者接近造型时,就增加原始造型的面数;反之,就减少原始造型的面数。这种控制保证了任

何时候原始造型的外观都保持平滑。

原始几何节点能被用来创建多种单一造型,但它们不能适应构造复杂形体的需要,VRML 中的 PointSet、IndexedLineSet、IndexedFaceSet 节点能使编程者用最基本的几何元素:点、线、面去构造精细逼真的形体,例如用许多微小的面可以近似地模拟出平滑的表面。IndexedFaceSet 节点中一个复杂的面有以下两种形式:

凹面:具有插入物或凹陷的面;凸面:不具有插入物或凹陷的面。

在缺省状态下,VRML 浏览器所创建的面都是凸面。为了创建一个凸面,可把 IndexedFaceSet 节点的 Convex 域值设为 FALSE,VRML 浏览器可将在这个节点内创建的凸面转换成多个凸面,通常为三角形。图 41.4-3 显示了人脸的侧面剪影分为多个三角形的情况。

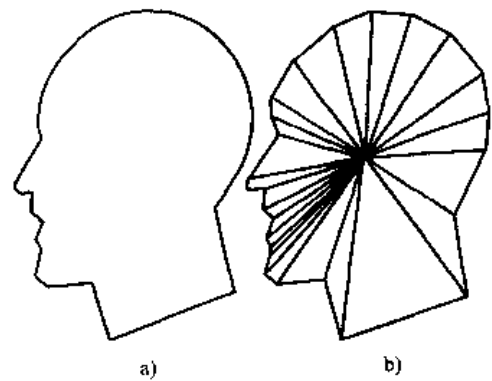


图 41.4-3 凸面的三角形表示

a) 人脸的凸造型 b) 被分割为多个三角形的人脸

在缺省状态下,VRML 浏览器假设一个面集中的所有面组成一个或多个实体造型,即这些造型的内部是看不见的,从而可以不绘制几何面的背面,大大节省绘制时间。这种优化处理称为 VRML 背面去除。但有时所需创建的造型并不是一个实体,如一个打开的盒

了,当观察者向盒子内看去,组成盒子所有面的反面必须画出。这时,可以通过设定 IndexedFaceSet 节点的 Solid 域值来允许或禁止背面去除。

在 VR 环境中,一个通常的三维造型是一个像山脉一样的崎岖不平的区域,它可以使用 IndexedFaceSet 节点的与之相接近的立方体表面栅格建立。对于栅格中的每一个表面, $X$  和  $Z$  的坐标值表明了区域表面的顶点, $Y$  坐标指定了表面顶点的高度或海拔。通过使用栅格中的大量表面,可以创建非常复杂的区域。由于这类栅格在造型中经常使用,VRML 特地指定了一个 ElevationGrid 节点来创立区域栅格。该节点可以在每个栅格点上指定区域栅格的定义和海拔,每个栅格立方体根据设定值形成区域表面,从而得到崎岖不平的区域。

在通常情况下,面集中的面被用来创建曲面的造型,如汽车的流线型车身。每个面都创建了一个小的平坦区域,如果用以构成造型的面较少,只能得到极其粗糙的近似造型。有两种方法可以避免表面的粗糙:一增加小平表面数来逼近造型的平滑表面;二将面以不同的方式绘制,但在视觉效果上似乎已使用多个面。第一种方法对造型精细程度提高的贡献是以牺牲绘制速度为代价的,在交互性具有更重要地位的前提下,它是不可取的。VRML 采用了平滑绘制的平面绘制技术,它在面与面边界的过渡过程中采用渐变的亮度,使得造型的面与面之间的界限含混不清。这样,实际构成造型的面数并没有改变,而造型的平滑程度大为提高。

有时造型的某部分并不需要进行平滑处理,如造型的尖锐部分。对于这类造型必须控制平滑绘制的过程,以便使造型尖锐的部分保持尖锐,平滑部分仍被平滑地绘制。

一个由面构成的造型其尖锐程度和平滑程度可以用两个相邻面的夹角来表示,该夹角称为折痕角。如果折痕角较小,这两个面就是物体上相对平滑的部分,应该进行平滑绘制;如果折痕角较大,这两个面就是物体上相对尖锐的部分,就不应该进行平滑绘制。图 41.4-4 显示了两个具有不同折痕角的平面。

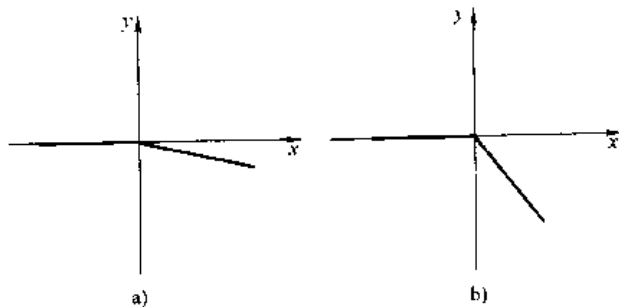


图 41.4-4 具有不同折痕角的面

a) 两个面具有较小折痕角 b) 两个面具有较大折痕角

为了控制平滑绘制,可以通过使用 IndexedFaceSet 或 ElevationGrid 节点的 CreaseAngle 域值设定一个折痕角的阈值,面集中任何一个面的折痕角如小于这个阈值就被平滑绘制,反之就保持尖锐。

#### (2) 编组节点及绘制的可见性选择

利用形状节点可以构造单个的简单或复杂形体,VRML 空间中的场景常常是包括多个三维造型的组群,在处理方式上需要作为一个整体对待。VRML 的几种编组节点实现了这一功能。它们可将任意数目的节点组织在一起。例如,可以创建一个建筑物的所有造型放在一个编组节点中,然后把建筑物作为一个整体来平移、旋转和缩放,并将它放置在一个城市的街区中。组成街区的建筑物也可以编成一个组并放置在一个城市中。利用 VRML 的这种组织机构,编程者就能操纵一个巨大的虚拟空间。

编组节点分为 4 类:基本组、转换组、变换组和布告板组。

1) 基本组 (Group) Group 节点提供了基本节点编组特性,使用 Group 节点的 Children 域,可以组织任意数目的节点。由 Group 节点编成的组作为一个整体,可以用 DEF 来定义,并在 VRML 文件中用 USE 重复引用。

2) 转换组 (Switch) Switch 编组节点扩充了 Group 节点的基本编组能力,它使用组的子节点组成一个选项表,一次从中选择一个来创建。它能将一个造型的不同版本组织在一起,作为 choice 域的域值。通过使用 Switch 节点的 whichChoice 域,可以选择待创建组中的子节点,这样不需对 VRML 文件作太多修改就可以在不同造型之间迅速转换。

3) 变换组 (Transform) 群组的平移、旋转、缩放功能通过 Transform 节点实现。Transform 节点中的 Translation 域创建新的坐标系,然后相对于父坐标系平移。当坐标系被定位后,该坐标系中的造型也被定位。Rotation 域将新坐标系相对于父坐标系进行旋转,面 scale 域通过设定  $X$ 、 $Y$ 、 $Z$  三个方向上的缩放因子改变原始造型的三维尺寸。

4) 布告板组 (Billboard) VRML 中的 Billboard 节点创建了一个特殊用途的布告板坐标系。布告板坐标系自动地旋转,将组中的造型作为一个整体,始终面向观察者。这个节点允许编程者制作出始终面向观众的布告板,用来作为提示符、帮助信息、广告、注解、控制面板和状态显示。

为了实时创建虚拟空间,VRML 浏览器必须检测界面与视野位置的关系,进行可见性选择。在视野内的造型必须绘制,在观察者背后的造型则被跳过。为了加速可见性选择,浏览器围绕 Group、Transform 和

Billboard 组的造型创建一个不可见的约束长方体 (Bounding Boxes)。该长方体的大小恰好能够将组的造型包括在内。在绘制时,浏览器并不检测一个组中每个造型的可视性,而是首先检测约束长方体的可视性。如果约束长方体是不可见的,那么约束长方体中的造型都不可见,浏览器不必再单个检测造型是否可见。如果约束长方体可见,那么浏览器再检测组中每一个造型的可见性。

### (3) 传感器节点与 VRML 的交互性

虚拟现实的最大特点是它具有人机交互的功能,用户能随心所欲地操纵、控制 VRML 环境中的物体。传感器便是 VRML 中提供交互能力和状态行为的单元。它具体包括 Cylinder Sensor, Plane Sensor, Sphere Sensor, Time Sensor, Touch Sensor, Visibility Sensor, Proximity Sensor, Collision 这八个节点。它们提供用户与虚拟世界中的物体进行交互的机制。从具体功能上来说,传感器节点使用一个定点设备(如鼠标)来感知用户的动作并产生相应的事件(event)。这一事件经过路由(route)将消息传送到相关节点,从而使虚拟世界中的对象发生响应。所有传感器都不产生视觉表达,只负责产生事件。

Touch Sensor 是一种用来检测观察者的接触并将事件输出以触发动画的传感器,它的输出说明了在何时、何地观察者接触到了可感知的造型。应用程序都可以根据定点设备(如鼠标)在屏幕上的移动(包括点击、拖动)来产生指定的事件。典型的如 Windows 消息对话框的弹出。Touch Sensor 节点可以添加到任何一组造型中,感知观察者对组中造型的任何一部分所做的鼠标移动,点击,拖动各动作。一旦检测到观察者与造型发生交互,便通过相应的 is Over, is Active 或 Touch Time 等领域的 Event out 事件输出一个布尔逻辑值 TRUE,该值被路由到其他节点的 Event In 事件,从而触发虚拟空间中的一个动画。当观察者与物体的交互结束,Touch Sensor 节点输出一个 FALSE 值路由到相应节点以停止动画。VR 环境中的物体能够按一定物理规律被用户操纵,利用 Plane Sensor, CylinderSensor 和 SphereSensor 三个节点可实现这一功能。它们能感知用户鼠标点击或拖动几何物体,根据用户意图使物体产生相应的移动。具体来说:

- 1) Plane Sensor 在局部的 XY 平面输出一个转换,引起造型平移。
- 2) Cylinder Sensor 输出绕本地 Y 轴一旋转值,感觉象旋转棒子盘。
- 3) Sphere Sensor 输出一个旋转值,感觉如同一个虚拟轨迹球。

当把这三种节点附于适当的物体上,便能使用户

对该物体进行平移旋转等操作。它们尤其适用于创建 VR 场系中的交互式物体如可开关的门窗、杠杆或滑动装置,可移动的日常用品等。

以上 3 种节点只是用以检测用户的动作,VRML2.0 还提供了三个感知用户存在的传感器节点: Visibility Sensor, Proximity 和 Collision。它们的存在体现了用户与计算机之间的真正交互;不仅用户能操纵 VR 环境的物体,虚拟场景也能以一种主动的方式感知用户的存在。

Visibility Sensor 节点通过构造一个视锥体来判断观察者是否存在于一个特定的视野区域内; Proximity Sensor 节点定义了一个长方体区域的检测空间来感知观察者是否接近某一物体; Collision 节点检测空间中观察者与造型接近和碰撞的时间。利用这 3 个节点的输出事件可以启动和停止动画及观察者与碰撞的声音。

### (4) 内插器节点与关键帧动画

从概念上来说,一个动画可以看成是有一个触发和逻辑单元,一个时间传感器,一组内插器及受时间传感器驱动的对象。在 VRML 中内插器节点与时间传感器节点为动画的实现提供了基础。TimeSensor 节点可以用来创建一个时钟,它提供了开始动画,结束动画和控制动画播放速度的特性。随着时间的流逝,这个传感器会产生事件来表示时间的变化。为了描述动画过程中发生的变化,VRML 提供了 PositionInterpolator, OrientationInterpolator, ColorInterpolator 和 ScalarInterpolator 四种内插器节点,分别用来控制 VR 环境中物体的位置和缩放、方向、颜色和透明度变化。这些内插器节点根据从时钟得到的信息,从相应的索引表中获得适当的一组值。在动画过程中,这组值被输出到对应节点的域从而决定了新的物体状态。

利用这些节点来建立一个动画路线,可以将 Time Sensor 节点的输出绑定到内插器节点的输入上,然后将内插器节点坐标  $x^*$ 、 $y^*$ 、 $z^*$  的变换矩阵表示为点的输出连接到 Transform 或 Material 节点的输入。一旦连通,时钟就开始计时,事件便在上述路线中路由,从而在动画回放中,节点的坐标系发生平移、旋转和缩放,造型的外观属性也发生变化。

一个动画描述必须提供物体在每一个新时刻的状态值,最直接的方法是使用一个列表将物体在所有的状态值都描述出来,但这种可能的状态值是无限的,要作这样一个列表是不可能的。VRML 中的动画描述使用了关键帧技术,它只需提供物体在几个关键时刻的状态值作为动画的框架,其中间值由内插器节点在关键帧之间进行线性插值得到。动画的回放速度决定了在关键帧之间插值数目的要求,对于一个快速动画,在

动画的启动和停止之间, VRML 只需绘制几次, 这种情况只需计算少量内插值; 而一个慢速动画则要求多个内插值来满足在一个较长时间段内 VRML 空间的多次绘制。

#### (5) 脚本节点和交互式动画

VRML 提供的感应器和内插器可以使造型产生相对规则的动画, 但对于复杂的动画就显得比较局限。VRML 提供了结合关键帧动画技术与交互式输入的可能性。这样既能由一个简单的用户输入触发已制作好的动画, 又能连接层次化和顺序化动画, 实时仿真, 甚至可以让动画角色自动进行复杂的动作。所有这些技术要求使用一个脚本节点。脚本节点可以看作是一个节点的外壳, 它本身没有动作, 创作者需要通过程序脚本来赋予脚本节点动作。程序脚本实质上是一种简化了的应用程序, 它要通过编程语言 ECMAScript 或者 JavaScript 编写。利用程序脚本和脚本节点, 可以创建一些节点让它们产生复杂的动作。

脚本节点可以描述一个编程人员自定义的传感器和内插器, 这些传感器和内插器需要一些关于域、event In 事件和 event Out 事件的列表及处理这些域和事件时所要做的动作的描述。一个完整的脚本节点的定义如下:

```
Script {
  url [ ] #exposedfield MFString
  must Evaluate FALSE #field SFBool
  directOutput FALSE #field SFBool
  #Any number of
  Field Field Type Fuel Name initial Value
  event In event In Type event In Name
  event Out event Out Type event Out Name
}
```

其中 url 域的域值是已编写好的程序脚本文件, 节点定义“#Any numlepor”后面的部分是编程者定义的域和输入输出接口, 下面的一个计算物体, 匀加速运动轨迹的 VRML 程序显示了为脚本节点定义的接口。

```
Script {
  Field SFFloat acceleration 2.0
  Field SF Vec 3f velocity 1.5 0.0 0.0
  Field SF Vec 3f position 0.0 0.0 0.0
  event In SF Vec 3f set-position
  event Out SF Vec 3f position -changed
  event In SF Float set..Fraction
}
```

程序脚本可以对以下三种类型的动作作出响应:

初始化

程序中止

事件接收

初始化发生在一个程序脚本没有接收过任何事件并且在其被第一次读入时。这是一个程序脚本将来接收事件做准备工作的时机。程序中止发生在程序脚本被中止时, 它是一个程序脚本将它所残留的东西清除的时机。事件接收发生在一个新的输入事件被程序的 event In 接收时, 此时程序脚本接收新的事件值, 并对此值进行计算, 产生一个或多个事件, 通过 event Out 事件接口发出。

从概念上讲, 当程序脚本每通过其接口接收到一个值后, 脚本节点就执行求值, 如果在一个 VRML 程序中包含许多脚本节点、感应器和内插器, 那么系统就会由于每次变化对每个节点进行的求值而经常性地陷入繁忙, 画面更新的速度减慢, 使虚拟世界的实时性和交互性变差。为了防止出现这种情况, 可以设置脚本节点的 Must Evaluate 域值来延迟对脚本节点的程序脚本的求值, 直到一个相对空闲的时间到来。在大多数情况下, 进行延迟求值不会给动画造成太大影响。但对于有一种情况, 观察者的动作必须被及时送到系统进行处理, 如在游戏中是否开了一枪, 是否跌入陷阱, 如果延迟求值就会拖延这些信息的发送, 从而使虚拟系统作出错误的响应。因此, 对于这种情况, 即使牺牲一些画面更新速度和交互性也要保证对程序脚本的实时求值。

从上面的分析中可以看出, Script 节点赋予了创造虚拟世界更大的自由度, 为 VRML 提供了更强大功能和集成能力, 创作人员可以通过编写程序脚本实现复杂逼真的动作效果, 使 VR 环境更具有沉浸性和真实感。

### 1.2.3 VRML 的细节层次控制

虚拟现实中的三维物体是若干个极微小的多边形组合而成, 对物体的绘制实际上就是对组成它的这些多边形进行着色渲染。一个复杂的造型往往由几万甚至几十万个多边形组成, 如果还对这个物体表面贴纹理, 那么着色绘制时的计算量将极其庞大。如果虚拟环境中的这类物体不在少数, 则系统将要花费一定的时间去创建它们。为了保证虚拟世界的交互性, VRML 采取了细节层次控制技术。这种技术实际上是在虚拟环境的真实感和交互性之间取得一个平衡, 通过损失一定真实感来换取对我们来说更为重要的交互性。

当观察者在虚拟世界中运动时, 他对远距离的造型细节的分辨率很低, 因而远处造型即使细节层次降低也不会影响整个环境的可信度, 而造型细节的减少使系统的绘制时间减少, 从而使交互性提高。基于这一

原理,VRML 提供了 LOD (Level of Detail) 节点来控制造型细节层次。一个复杂的造型可在多重版本下创建,一般分为高级 (high)、中级 (middle) 和低级 (low) 三种细节。高级细节版本具备所有真实的细节,

中级细节版本具有造型的基本组成部分,低级细节版本具有更少的成分,如去掉纹理,只保留物体的大致形状。图 41-4-5 是一个造型的三个不同细节层次的版本,其细节层次由多边形数目决定。



图 41-4-5 不同层次细节的人头模型

LOD 节点可将描绘不同细节层次的 VRML 节点组织在一起,系统根据用户视点的位置计算出与 LOD 节点中造型中心点之间的距离,并通过一定值判断距离的远近可自动切换造型的不同版本。不只是范围值可以作为造型版本转换的线索,其他的执行条件也可以控制版本的转换。例如,如果用户试图在虚拟空间中尽可能保持交互式运动,则可以在观察者运动时选择造型的低细节版本,而在观察者停止运动时转换到高细节版本。

可以看出,VRML 的细节层次控制优化了绘制速度,这在一定程度上提高了 VR 系统的实时性和交互性。

#### 1.2.4 VRML 的纹理映射及控制

在创建 VR 环境造型后,为了增强其真实感和可信度,往往需要赋予造型以适当的纹理,这种技术被称为纹理映射。纹理可从实际照片中提取或人工产生,可以是全透明或半透明的,并支持真彩色,纹理图像可以存储为彩色图像或灰度图像,彩色图像存储图像中的每一像素的三色及透明度的值,灰度图像只存储每一个像素的灰度等级。使用彩色图像时,纹理图像的颜色被映射为 VRML 造型的一部分,通过指定材质节点的相关域值,可以生成半透明的或发光的纹理造型。使用灰度等级图像时,灰度值作为加权因子与材质节点的颜色域值作用,可以逐点地加亮或变暗造型的颜色。

目前,VRML 支持四种最常用的图像文件格式: JPEG、GIF、PNG 和 MPEG 其中,前三种图像文件格式用于存储单个的静态纹理图像,最后的 MPEG 格式可存储动态图像。大多数彩色图像文件格式以三个 8-bit 整型值来存储 R、G、B 三色值,每个 8-bit 整型值在 0 到 255 之间,VRML 自动将这三个整型值映射为 0.0 到 1.0 之间的浮点 RGB 颜色值,其转换过程为将三色值除以 255。

VRML 的纹理映射是采用 MIP 贴图 (MIP MAPPING) 技术中的三线性 MIP 纹理映射 (TRI LINEAR MIP MAPPING),它所利用的三线性插值是最高质量的纹理函数之一,其算法是先在包含被纹理映射的多边形的两个 MIP 图像之间进行两次双线性插值,然后将每个被纹理映射像素的九个调配项的插值结果调配在一起,这样生成的图像即使在非常锐利的衔接处也相当稳定,当纹理相对视点有移动时,没有显著的过渡段。VRML 还用到了四线性三维 MIP 纹理映射方式,它是一个有效的三维纹理的三线性插值,能自动生成一系列的三维体,每个三维体比它上面的三维体小八分之一,系统两个相邻体之间的 MIP 图中的八个相邻像素之间插值,然后将这两个结果调配在一起以达到四路插值。

#### 1.2.5 VRML 的光照和雾化

VRML 中的光源与现实世界中的光源有着相同的目的,照亮一个场景并且突出关注点,光源的特性决定入射光的方向、强度及颜色,物体几何顶点性质及其表面材料决定反射光的方向、强度及颜色,VRML 只在几何顶点上执行光照运算,简化了在所绘图的每个像素上执行光照运算的复杂处理。

在缺省情况下,VRML 提供 1 个头灯,该光源跟视点一齐运动,始终照亮视点的正前方,VRML 还支持其他三种类型的光源:点光源、平行光源和聚光光源,点光源是呈放射状向四周发出光线的光源,由其辐射出的光线似乎来自一个点,在 VRML 中,通过指定其在空间中的三维位置、亮度和颜色来生成一个点光源,平行光源是无限远光源,发出的光线平行指向同一方向,在定义平行光源时除了要规定那些点光源所具有的性质外,还要指定光源的照射方向,聚光光源放置在 VRML 世界中的某一点,照向一特定的方向,所发出的光线集中在一个锥体内,只有在该圆锥体中的透

型才能被照亮。

在VR环境中可以采用多个光源,它们的类型、位置、颜色可以不同,为了避免造型被不希望的光源照射到。三种光源节点提供了相应的控制手段:点光源节点中的radius域规定了它作用的最大半径,通过设定适当的域值可使光源照亮需要光照的造型;聚光灯的光照锥体参数可调,因而可限制光照范围;平行光源对物体的照射要将平行光源节点和物体造型节点编组成一个Transform或Group节点才能发生作用,所以对于需要平行光源照射的物体可以采用编组节点进行控制。

点光源和聚光光源均存在光线衰减(Attenuation)问题。随着光源与表面距离的增加,光源对表面的影响会减小。VRML衰减操作的目的是模拟这种实际效果。衰减是光源与它所照射表面之间距离的函数。VRML提供三种衰减因子:constant,linear和distance-squared,衰减系数的定义为:

$$F_s = \frac{1}{K_0 + K_1 d + K_2 d^2}$$

式中  $d$ ——表面与光源之间的距离;  
 $K_0, K_1, K_2$ ——常数、线性和二次型衰减因子。

VRML使用雾(Fog)控制空间中的大气,虚拟雾有两个属性:颜色和浓淡。雾的颜色缺省值为WHITE,也可选择别的颜色以产生相应的环境气氛。Fog节点中的Visibility Range域控制雾的浓淡,同时也可以通过指定雾的类别控制浓淡。VRML中有两种类型的雾,线性雾和指数雾。线性雾的浓度随着与观察者视点距离的增加而线性增大,适用于产生薄雾。指数雾的浓度随着距离的增加而呈指数增长,能生成更为真实的浓雾。

在现实世界中,雾中的小水滴悬浮在空气中遮挡视线,影响我们观察物体的边缘和颜色。对于这种现实雾的模型,应当计算虚拟大气中每一个小水滴的效果,但是它需的计算量不是VR系统所能接受的。VRML采用简化的方法模拟雾的效果,通过判断造型与观察者的距离决定造型的外在颜色。距离观察者较远,已在雾中的造型被遮掩,变成和雾一样的颜色,而距离较近的造型保持了它们本来的大部分颜色,受雾的影响较小。

VRML中的环境渲染在虚拟环境的交互性和实时性得以满足的前提下,通过极节省的方法实现了真实世界的环境效果,并且其模拟质量在对环境细节要求不高的场合下是令人满意的。

### 1.2.6 VRML的背景及添加声音

在建立一个虚拟世界时,可以通过使用各种节点

生成观察者四周的物体,如地板、屋顶、墙、家具等等。你可以建立围墙或曲折的门厅以限制观察者的视野。采用这些方法可以减少在观察者移动时VRML浏览器所必须重画的造型的数目。这使得观察者在观察你建立的VRML世界时,VRML浏览器减少描绘的时间,提高翻译速度。

但是,在某些情况下,让观察者脱离封闭的、有限的世界是必须的。观察者就可以看到你VRML世界中的天空和地平线,你可以采用三维造型构造天空,并在地平线处放置山脉或建筑。但是,若观察者从不飞上天空或在地平线处来回走动,建立这些三维世界将浪费浏览器的描画时间。

这时,可以利用VRML提供的背景节点。利用该节点你可以控制天空和地面的颜色,设置远方山脉和城市的景象。这使得你可以在不生成三维天空和远景造型的情况下为你的VRML世界提供一个外部环境。

由VRML的背景节点生成的背景由一个天空球体、一个在天空之内的地面球体、以及在地面球体之内的背景立方体组成。这三者在概念上均为无限大,可以包括整个的世界,观察者可以从不同的角度进行观察,但是不能接近它们。使用背景节点既可丰富VRML世界,又可极大地节省浏览器的描画时间。

在生成虚拟世界后,为了使其更真实,可以增加声音。你可以生成环境声音,例如在虚拟的林荫路中的背景音乐等。你也可以利用路线触发声音,例如一扇门关闭时发出砰的声音,小鸟发出啾啾的声音等。VRML中声音可以进行定位,例如左侧的一扇门关闭时,可以控制使得只有左侧的扬声器或耳机中发出相应的声音。

VRML中声音涉及两个部分:声源和声音发射器。声源和声音发射器与家用音响系统的功能类似。声源提供声音信号,例如CD唱机或录像机。声音发射器将声音信号转换为你可以听到的声音,就像音响系统中的扬声器。

VRML中有两种类型的声音文件可作为声源:①数字声音文件;在其中存储数字化的记录的声音,例如音乐和音响效果。VRML提供一种十分普遍的存储数字声音的方法,即WAV文件格式;②通用MIDI类型I文件:包含如何演奏一段音乐的指令。通用MIDI是在PC和MACINTOSH上演奏音乐MIDI一个特定变化。

VRML中可以有任意数目的声音节点。一般情况下,有一些声音节点用于播放背景音乐和噪声,其他的声音节点则由路线触发,导致播放一些音响效果。你也可以利用范围椭球,限制每一椭球的效果,使各种音响效果不被同时听到。

在 VRML 中可以轻松控制声音的音高、范围、优先级、是否可重放等。声音节点的利用极大地增添了虚拟世界的魅力。

### 1.2.7 VRML 的视点控制

在虚拟空间中,我们可以来回移动,并从不同的位置来观察它。在每一个新的观察点浏览器获得的图像,就像是在使用照相机在你的屏幕上播放一样。你在虚拟空间的不断移动,使得照相机不断地调整其位置和朝向。

在 VRML 虚拟空间中可以很方便的建立预定的视点。视点就是你在 VRML 中预先定义的位置和朝向,就像是推荐给旅游者的观光地点。我们可以预先设定很多视点,观察者可以从浏览器菜单中选择不同的视点,相当于旅游者在不同的地点开始观光,这样增强了观察者和虚拟空间的交互性。

为增强临场感,在虚拟现实技术中使用替身象征性的代替现实世界中人。使用替身技术,真实世界中的人可以在虚拟世界中穿行,他看到什么,你就能看到什么。替身在虚拟世界中穿行,要有移动的方法。VRML 中通过定义替身的导航信息来规定其移动方法。例如有飞行、行走等方式。

灵活的运用视点和替身技术将极大增强 VRML 虚拟世界的吸引力,而且也是大型 VRML 空间的必须技术。视点的关键是定义好位置和朝向。定义适当的视点数目也是相当重要的。替身的关键是他的尺寸、速度和能见度范围等。

### 1.2.8 VRML 中程序脚本的灵活应用

我们知道虚拟世界中的动画节点多种多样,例如: TimeSensor 节点、各种内插器节点、感知观察者接触和拖动的节点以及感知观察者接近和碰撞的节点。通过这些节点,你可以使虚拟世界中的三维造型运动起来。但是,这些节点只能产生相对简单的动画动作。即需要询问场景中的状态,然后根据这个状态做出决定。也即只能静态改变场景的部分内容。要实现控制“门的开和关、灯的灭和亮”或更复杂的逻辑块这样简单的场景将是十分困难的。

要实现这种复杂的逻辑块,则 VRML 必须具有响应动态行为的能力。这种动态行为与静态行为有着极大的不同,这是因为它能够实现比按原来定义的路线更多的路径事件。动态行为需要能力来询问状态,基于这些状态做出决定,再在这些决定的基础上改变场景的状态。在 VRML 中并没有预先定义的节点来做这项工作,这些功能的实现要用到程序脚本。其中 SCRIPT 节点很有作用。

例如,利用 SCRIPT 节点构造一个布尔事件过滤器:

```
DEF Filter Script {
  url "javascript:
  function set_boolean(bool,eventTime){
    if(bool==false) {return;}
    if(value==true) {value=false;}
    else           {value=true;}
    value_changed=value;
  }"
  field SFBool value FALSE
  eventIn SFBool set_boolean
  eventOut value_changed
}
```

利用这个过滤器,你可以很容易的实现“门的开和关、灯的灭和亮”。因为该过滤器能够像一个开关一样,第一次点击就发送一个 TRUE 值,第二次就发送一个 FALSE 值,第三次又变为 TRUE 值,依次类推。

总之,程序脚本可以完成如下操作,诸如初始化、程序终止、事件传送、数据类型转换和在程序设计语言的 API 中定义浏览器访问。Java 和 Java Script 语言都提供了这些环境,有了它程序脚本就可以实现这些操作。灵活运用程序脚本可使你的虚拟世界充满魅力,程序脚本的存在也使得 VRML 有了更强的生命力和号召力。

## 1.3 VRML 的应用及常用资源

VRML 在近年发展迅速,大量的成功的应用出现,其应用几乎遍及工业的各个领域。VRML 在远程教育、科学计算可视化、工程技术、建筑、电子商务、交互式娱乐、艺术等领域都有着广泛的应用前景,利用它可以创建多媒体通信、分布式虚拟现实、设计协作系统、电子商务、网络游戏、虚拟社区、虚拟空间等全新的应用系统。

### 1.3.1 典型的应用领域

#### (1) 教育应用

把 VRML 用于建造人体模型、电脑太空旅行、化合物分子结构显示等领域。由于数据更加逼真,大大提高了人们的想像力、激发了受教育者的学习兴趣,学习效果十分显著。随着计算机技术、心理学、教育学等多种学科的相互结合、促进和发展,能够提供更加协调的人机对话途径。华盛顿大学外语系 Carlos Seligo 说:“外语系设想通过远程教学授予学位,但一直苦于找不到具有交互能力的虚拟教室。我认为 VRML 将成为虚拟教室的黑板,而教师们知道该使用什么样的粉笔。”

### (2) 工程应用

当前的工程很大程度上要依赖于图形工具,以便直观地显示各种产品,目前CAD/CAM已经成为机械、建筑等领域必不可少的软件工具。VRML将使工程人员能通过因特网或内部网按协作方式进行三维模型的设计、交流和发布,从而进一步提高生产率并削减成本。

### (3) 商业应用

对于那些期望与客户建立直接联系的公司,尤其是那些在他们的主页上向客户发送电子广告的公司,Internet具有特别的吸引力。VRML有可能大幅度改善顾客购买商品的经历。例如,顾客可以访问虚拟世界中的商店,在那里挑选商品,然后通过Internet办理付款手续,商店则及时把商品送到顾客手中。

### (4) 娱乐应用

娱乐领域是VRML的一个重要应用领域。VRML是一个功能十分强大的在Internet网上交换三维场景的描述性语言,它能提供更多的多人之间的交互,提供更加逼真的虚拟环境,从而使人们能够享受其中的乐趣,带来更好的娱乐感觉。VRML目前正朝着实时通信、大规模用户交互的方向发展。

## 1.3.2 成功案例

### (1) 科技探索

一个对火星好奇的学生登上虚拟的宇宙飞船,从地面点火起飞,并尝试各种飞行路径飞向这个红色星球。在火星表面登陆后驱车绕着科学上正确的轨迹运动……带有合适的重力和光照。这一景象在火星探路者登上火星时已经借助VRML技术发生在全球Internet用户的眼前;利用三维的虚拟太阳系模型和来自火星的大量数据,物理和大文学世界被带入生活中,而且比任何书本或第一代Web页面更为吸引人,并具有交互性。

### (2) 商业宣传

百事可乐公司在自己的网站上创建一个广告画面:探路者在火星表面走动且发现了一瓶百事可乐!所有文件只有12KB,因此能十分方便地配置在一个普通的Web站点上。美国Construct公司完成了一个VRML项目,目的是展示VRML作为一个故事片媒介。片长11min,高分辨率逼真图形的故事占了不到1MB磁盘空间。相似长度和质量的一个MPEG电影将比它大300倍。

### (3) 电子购物

珠宝加工商为远距离客户加工珠宝,为了确保最终样式符合客户的期望,可应用VRML创作工具建立一个项链模型,然后把设计通过Web传送给顾客。虽

然客户没有拿到真正的东西,但仍能从所有角度考察这种项链,检查反射面的形状,以决定作为基片的宝石高度。因此,在设计阶段就把客户的需求呈现出来。类似的VRML应用范围很广,从医疗构图、分子造型、工程与设计、建筑到包含更多娱乐的游戏、各种广告、虚拟剧场以及日常生活的习俗,如起居室的选择和家具摆设、周末公园旅游计划、汽化器维修等。

### (4) 文物保护

意大利教堂的壁画被破坏已经很长时间了,但是仍有足够的信息使它恢复原来的模样。采用建筑造型恢复教堂和采用图像扫描技术恢复壁画,然后把它们集成到VRML文件中,这样,艺术学院的学员就能在走廊里行走,观看头顶的壁画在一天中的不同时候,随着光线的不同而呈现的不同图案,当然还可以增加鉴赏评论等有关壁画的信息。这种需求在我国更多,也更为紧迫,如敦煌石窟、兵马俑、长城以及众多的园林。

### (5) 娱乐业

Bitom是一家位于美国洛杉矶的娱乐公司,最近与Cosmo软件公司演示了第一部VRML动画片。

## 1.3.3 VRML 浏览器

VRML浏览器在VRML97标准发布后,曾经繁荣一时,其中以SGI公司的CosmoPlayer最为著名。99年后,由下一代VRML标准迟迟未能制订出台,各厂家纷纷推出附带扩展功能的VRML浏览器,甚至是不兼容VRML的浏览器,再加上SGI公司的退出,VRML陷入了停滞不前的境地。幸运的是,目前又有一批新锐公司如Blaxxun、Parallelgraphics正在活跃起来,强力推动着VRML大步向前,它们所出的浏览器Blaxxun Contact和Cortona在性能上已经超越了CosmoPlayer。以下是2000年1月份的VRML浏览器全球使用情况调查表,见表41.4-1,也许结果并不十分准确,但大致反映了目前的现状。

表 41.4-1 VRML 浏览器使用比率表

使用的VRML浏览器	使用者比率(%)
CosmoPlayer	40
blaxxun Contact	28
Cortona	28
WorldView	4
同时使用多种浏览器	32

### (1) SGI COSMO Player

该浏览器为SGI公司开发,目前最常用,目前最新版本是2.1.1。最后更新1998.12;应用平台Macintosh、SGI、Windows 3.x、Windows 9x、Windows NT。

主要特点:CosmoPlayer曾经是最好的VRML浏



览器,它出自VRML的早期领导者SGI公司,在速度、质量、兼容性等诸多方面,都曾遥遥领先,CosmoPlayer的控制面板至今仍是最好、最科学的,可方便地在WALK模式和EXAMINE模式之间进行切换,它的兼容性和质量至今仍是一流的。

#### (2) Blaxxon Contact 4.3

Blaxxon公司出品,最高版本4.3;最后更新2000.2;应用平台Windows 9x、Windows NT。

主要特点:Blaxxon已经取代Cosmo player成为VRML领域的领导者,它的浏览器Blaxxon Contact的最新版本在各项性能指标上,都十分出众,如全面支持VRML97,最先支持NURBS、UM;渲染速度名列第一,不足的是它的控制面板太难控制,浏览模式的切换十分不便。

#### (3) Cortona 2.2

Parallelgraphics公司开发,最高版本2.2;最后更新2000.7;应用平台Windows 9x、Windows NT。

主要特点:Parallelgraphics是目前VRML领域最有活力的一家公司,它的Cortona是Blaxxon Contact的最大竞争对手,除很好地支持VRML97、NURBS外,还支持多种自己规格的扩展功能,如键盘输入、拖放控制、Flash等,更是业内第一个(也是唯一)支持最新EAI功能的VRML浏览器,而它的安装文件仅有1.8M,是各种VRML浏览器中最小的,Cortona的兼容性做得稍差,主要应用环境是Windows平台和IE浏览器,并且声音功能还有一些小缺陷。

#### (4) WorldView 2.1

CA公司开发,最高版本2.1;最后更新1999.1;应用平台Windows 9x、Windows NT。

主要特点:IE5.0的内置VRML插件,性能落后,长期未更新,无使用必要。

#### (5) Community Place 2.0

SONY公司开发,最高版本2.0;最后更新1998.10;应用平台Windows 9x、Windows NT。

主要特点:主要应用于多用户环境,SONY虽然从SGI挖来了VRML领域的先锋者Chris Marrin等,但主要是进行Blendo项目的研究,SONY研究VRML的目的是为其游戏机PlayStation服务。

#### (6) Viscape 5.61

Superscape公司SVR浏览器插件,兼容VRML97;最高版本5.61;最后更新1999.2;应用平台Windows 9x、Windows NT。

主要特点:专用网络虚拟现实格式SVR的浏览器Viscape,在速度、质量、交互性等方面,比目前所有的其他VRML97浏览器都要优秀,对VRML97的支持不够。

### 1.3.4 VRML常用资源

1) VRML官方网站: <http://www.vrml.org>。

2) VRML Repository: 是有关最新的VRML软件、虚拟空间例子、文件和资料的主要Internet网站,该Repository有一个现在所有VRML浏览器列表和如何得到更进一步信息的方法,其URL是: <http://www.sdsu.edu/vrml/>。

## 2 实时视景开发工具——OpenGL Performer 基础

IRIS OpenGL Performer™是SGI公司开发的一个可扩展的高性能实时三维视景开发软件包,主要用于仿真可视化、娱乐、虚拟现实、视音频媒体以及计算机辅助设计等领域,目前已经发布了3.0, Performer提供了一组与标准C或者C++绑定的应用程序接口,它们可以运行于所有SGI图形工作站以及PC图形工作站上,它通过一个使用灵活的三维图形工具集来提供高性能的渲染能力。

### 2.1 OpenGL Performer 基础

#### 2.1.1 组成

SGI提供的Performer™是开发实时可视化仿真和其他图像应用程序强大的可扩展的编程接口,其同时支持IRIS图形库(IRIS GL)和工业标准的OpenGL图形库,SGI Performer软件包主要包括三个DSO(对IRIS、Linux系统)或者DLLs(对Windows系统),相应的支持文件和例程源代码组成,Performer是ANSI C/C++编写的类库,它包含有两个主要的库libpf和libpr和五个相关的库libpfda、libpfdh、libpfui和libpfutil、libply,如图41.4-6所示。

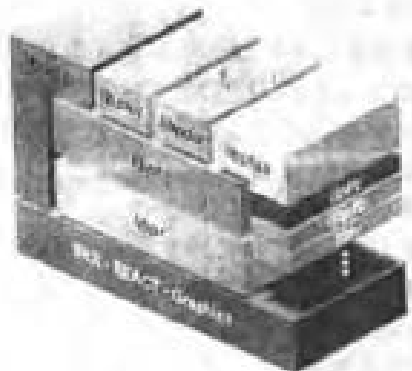


图 41.4-6 Performer 的类库结构

Performer早期只支持IRIS系统,现在已经支持Windows、Linux系统,未来会兼容所有操作系统,它

不将应用程序与操作系统和图形库相分离, 应用程序可以直接访问所有系统层, 它不仅包括 libpf 等 Performer 库, 还包括操作系统库。

#### (1) Libpr 库

Libpr 不面向多处理也不面向特殊应用, 它唯一的目的是使图形硬件的利用达到最优。

##### 1) 该库的四个基本元素为:

- 高性能的立即模式几何图形渲染。
- 高效的图形流水线 (Graphic Pipeline) 状态管理。
- 快速灵活的交叉点测试。
- 通用底层编程工具。

##### 2) 该库的一些重要特征为:

- 封装图形库状态。
- 层次受限。
- 不完全管理所有的图形库函数。
- 允许自由混合图形库及 libpr 调用。
- 提供底层图形库不具备的功能。
- 是一个简单的应用目标端口。

libpr 提供了对立即模式的几何图元 (如点、线、三角形和三角带 (Triangle Strips)) 的高性能渲染。渲染几何图形时, 几何图元的外观由几何流水线 (Geometry Pipeline) 的当前状态决定。高效地管理这些状态是取得良好性能的关键。为了达到最高的性能, 状态的变化必须最少, 渲染回路必须自主。为此, IRIS Performer 主要使用了两个机制: pfGeoSet 和 pfGeoState。

pfGeoSet 封装了三维几何图形, pfGeoState 封装了几何图形外观的参数。pfGeoSet 被绘制时, 它的顶点和其他属性被送往图形库处理, 最后在屏幕上光栅成像。pfGeoState 被应用时, 其封装的状态用于为一定的外观 (如光照、纹理和雾气) 配置图形库。pfGeoSet 经常引用 pfGeoState, 从而实现几何图形和外观属性的同时定义。当 pfGeoSet 被绘制时, 与它关联的 pfGeoState 被自动地首先应用, 以保证渲染的几何图形有合适的外观。

#### (2) Libpf 库

Libpf 在 libpr 之上, 主要提供高性能场景组织以及可扩展的多进程渲染体系。

#### (3) Libpfdm 库

数据库应用库 Libpfdm 定义三维物体的定义数据以及属性数据。

#### (4) Libpfdh 库

数据库 Libpfdh 包含了工业标准的各种三维定义格式输入包, 如 OpenFlight™, Open Inventor™, iv, OpenGL Optimizer™, csh, 等 70 多种。另外, OpenGL Performer 提供本地 PFB 格式, VRML 文件输入包目

前已经由的三方软件公司 Open World 公司提供。

#### (5) Libpfui 库

Libpfui 库包含了用户界面等操作部件, 用来定义人机接口。

#### (6) Libpfutil 库

Libpfutil 库是一个很有用的应用库, 提供了一系列的系统配置接口, 如多处理器配置, 多通道支持, 线程管理, 场景遍历函数等。

#### (7) Libpfv 库

模块化 libpfv 库提供了快速构建 Performer 的高层次应用结构, 通过简单定义 XML 的配置文件就可以定义显示配置、模块操作、漫游、拾取等, 是 Performer 最新的功能。

OpenGL Performer 从 3.0 版本包含了 libpfv 库, 它主要是为了建立一个模块化, 处理交互式应用的 C++ 库。主要有以下特征:

读写 XML 文件。

通过文件或者 API 调用来定义复杂的显示配置 (包括 pipes, windows, and channels)。

跟踪鼠标、键盘操作。

设置和三维场景元素的人机交互。

管理多场景。

光亮多视角。

程序模块化, 方便程序开发和管理。

上述具体库的描述, 参见 OpenGL Performer™ UserGuide。

### 2.1.2 渲染流程

Performer 提供多进程渲染结构, 它的内核通过多线程技术和并行渲染流水线技术实现对每帧视图的管理以及图形流水线输出图像的生成, 如图 41-4-7 所示。每一套软件渲染流水线被主要分成三个阶段:

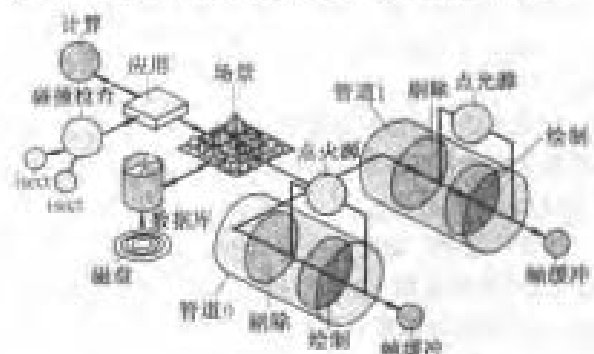


图 41-4-7 Performer 的渲染模型

- 应用阶段 (APP Stage), 定义视点种对象位置。
- 剔除阶段 (CULL Stage); 细节程度 (LOD), 可见性裁剪和消隐、排序。

• 渲染绘制阶段 (DRAW Stage): 渲染视景。

每一个绘制流水线使用一条图形流水线 (pipeline) 并能渲染多个输出通道 (channel)。用户可以自由配置 APP, CULL 以及 DRAW 三个阶段的任务配置, 如将三个阶段分配给单 CPU、多 CPU、指定 CPU、任务同步异步执行等。

### 2.1.3 节点和节点类型

OpenGL Performer 的应用场景结构是一个树结构, 如图 41.4-8 示。它允许共享实例, 共享实例提供了一个机制使得确保删除对象的安全操作, 树中的每

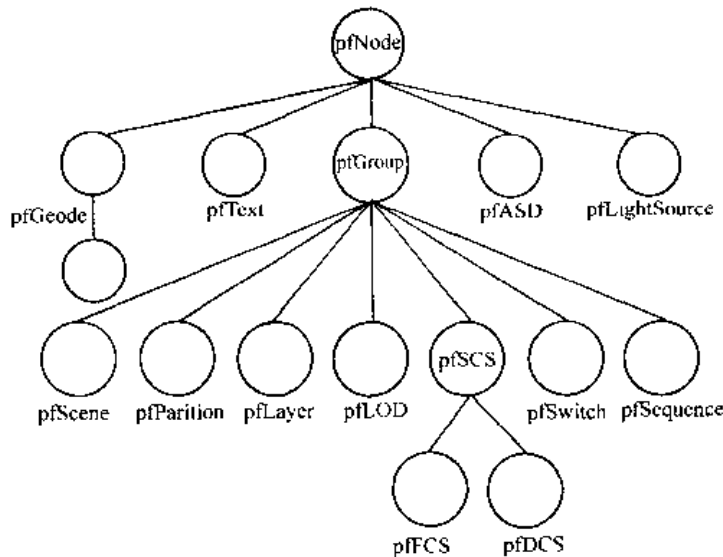


图 41.4-8 Performer 场景数据树

一个节点都是一个实体。典型的是场景树结构中的叶代表几何物体, 内部节点代表转换对象。场景树中的子树也可以代表场景中一个物体, 这个物体本身又是由一些子元素构成。如果物体的多个备份要组成一个场景, 那么在图形的不同地方应用同一棵子树是最高效的, 子树的任何变动都会在其他实例中反映出来。OpenGL Performer 为创建、排序、修改和遍历树状场景提供了很多有用的函数。

#### (1) 节点类型

所有节点类型可以归为三个基本类: 根节点、分支节点 (内部节点) 以及叶节点。节点类型的命名规则为: pf<名字>。

根节点是一副视景图像中最顶层的节点。遍历时引用根节点, 从而获得一幅完整的视景图像。pfScene 是 Performer 的根节点。

叶节点包含一副视景图像中所有的可视几何图形数据。这些数据包括: 几何图形的构成数据、几何图形的属性、绘制几何图形时图形子系统的状态元素。一个叶节点中的数据与整个图形子系统的状态数据相结合, 可以完全定义一个可视化对象的渲染构架。

分支节点包含遍历函数所需的控制信息和状态信息。分支节点可用来指明遍历的路径, 对几何图形施加变换。表 41.4-2 给出 Performer 的基本节点类。

表 41.4-2 Performer 的基本节点类

节点类型	节点类	描述
pfNode	抽象节点	基本节点类型
pfGroup	分支节点	0-n 个子节点形成的组
pfScene	根节点	可视化数据库的父类
pfSCS	分支节点	静态坐标系
pfDCS	分支节点	动态坐标系
pfSwitch	分支节点	转换
pfSequence	分支节点	子节点的顺序, 动画时使用
pfLOD	分支节点	细节程度节点
pfLayer	分支节点	渲染共面几何体
pfLightPoint	叶节点	定义点光源
pfLightSource	叶节点	定义光源
pfGeode	叶节点	定义几何元素
pfBillboard	叶节点	定义跟随视点旋转的节点
pfPartition	分支节点	划分几何体以提高遍历效率
pfMorph	分支节点	变形节点
pfText	叶节点	文字节点

#### (2) 属性继承

视景分层结构中最基本的元素就是节点, Performer 提供多种节点类型的同时, 使用了类继承的概念, 允许不同节点共享某些属性。

pfNode 是最基本的节点类, 所有节点都继承它的

属性。pNode 又是一个抽象类，它不能在应用程序中被创建，它的目的是为节点类型的层次结构提供一个概，并定义所有节点类型共有的属性和操作。这些属性包括：节点名、父节点的列表、几何图形的边界等。操作包括：新建、复制、删除和查找节点、获取和设置节点属性、打印节点数据。

(3) 使用节点

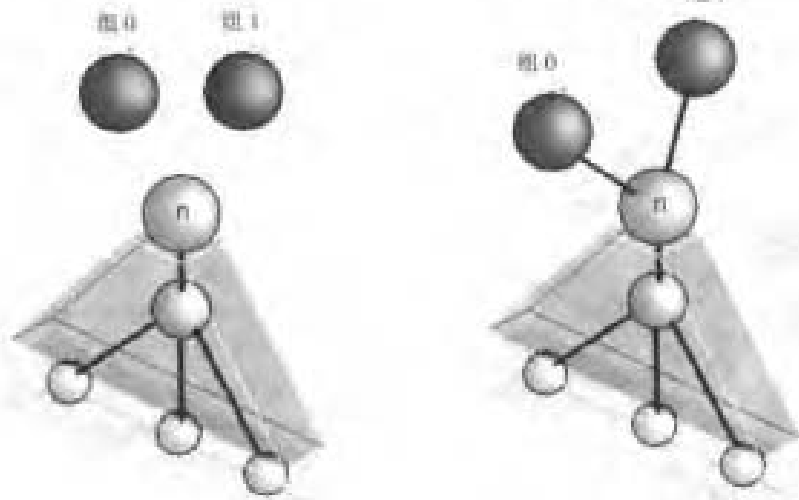


图 41-4-9 共享实例

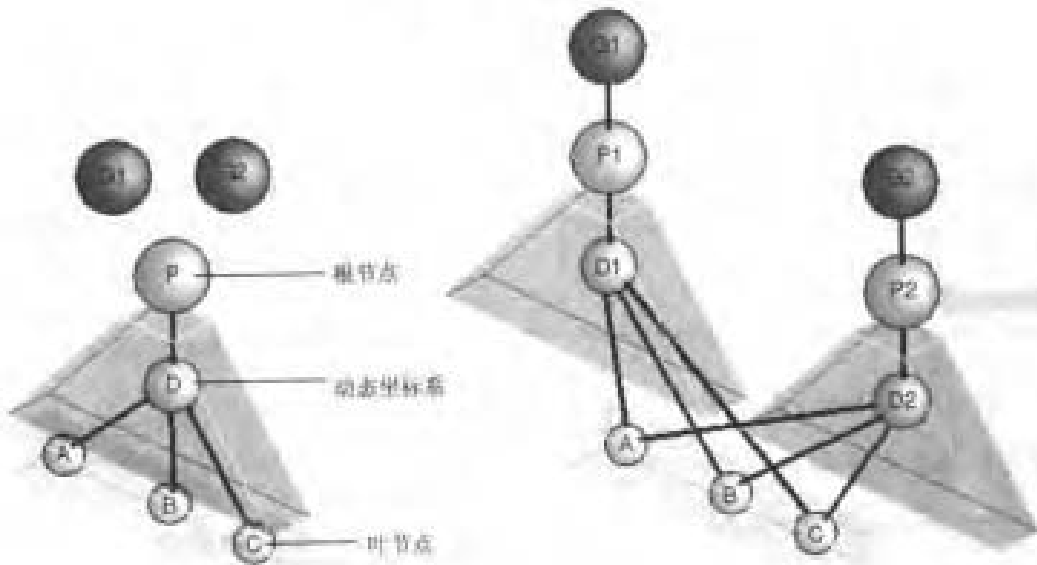


图 41-4-10 克隆实例

2.1.4 数据载入

OpenGL performer 为三维模型和三维环境提供了多层支持，主要有三个库，如表 41.4-3。

Libpdu 提供了一个方便的函数 pldLoadFile() 载入多种格式的数据库文件，其可以在当前路径搜寻文

Performer 中 libpf 支持两种实例化方法。

共享实例 (图 41.4-9)：简单地在父节点中加入一个指向该节点的指针，不复制任何数据。如果被实例化的节点有子节点，则子节点被实例化。其最有效地利用了资源，适用复制静态几何体。

克隆实例 (图 41.4-10)：使用克隆实例方法创建的节点共享叶节点。这种共享方法可以独立运动。

件，抽取扩展名，对扩展名进行转化，形成相应动态共享对象的实例，动态载入动态共享对象，形成相应的载入函数的名称，最后在动态共享对象库中发现该函数，调用函数。其是一个重载函数，读取其他格式，不需要改变应用程序，图 41.4-11 给出 OpenGL Performer 与多种数据库格式的关系。



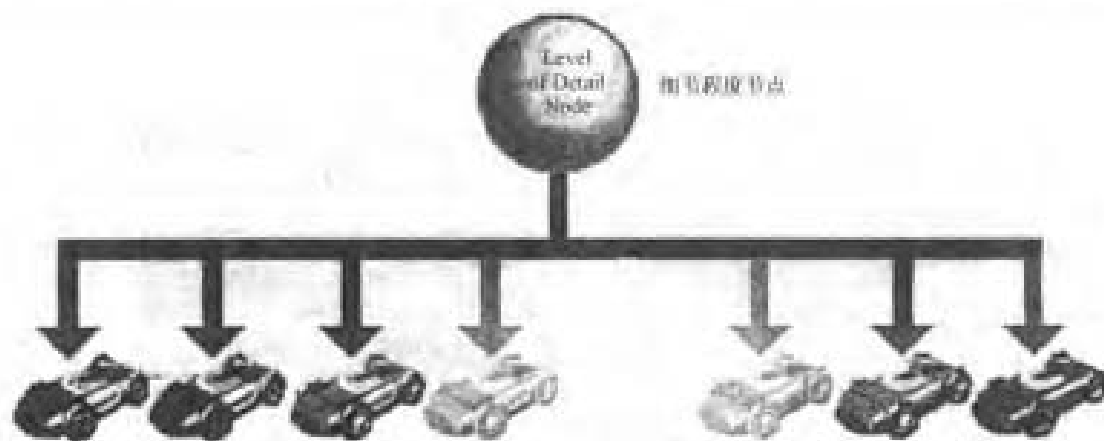


图 41-4-12 细节程度结构

2) 细节程度范围 细节程度的切换范围在用于细节程度选择之前被处理。设置切换范围的目的是为了当对象接近一定可见范围内,切换细节程度模型,对象的细节程度受通道大小,视野范围和观察者距显示对象表面的影响。

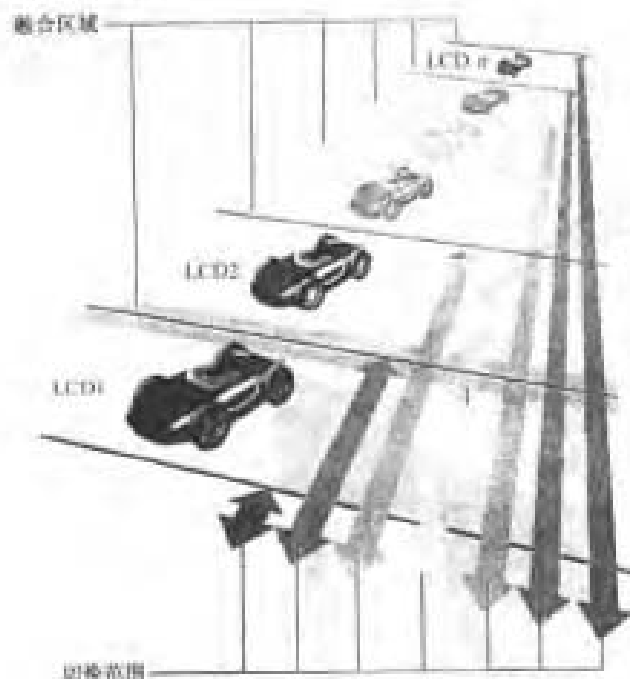


图 41-4-14 细节程度处理

#### · 细节程度的过渡

两个不同细节程度的融合过渡,如果太突然,显得生硬,需要对切换过程做处理。这种处理是对两个细节程度模型进行模型融合。如图 41-4-14,用水平阴影表示。

融合区内模型的绘制方法是利用透明性分别采样前后两个细节层次模型,并随着距离的变化,逐渐减少对前者的采样而增加对后者的采样,直至最后所有的采样都来自后者,如表 41-4-4 所示。在采样的同时将采样组合成新的模型,然后绘制。

表 41-4-4 细节程度的融合过渡

距离	LOD1	LOD2
融合区的前边缘	不透明	全透明
融合区的中部	半透明	半透明
融合区的后边缘	全透明	不透明

#### (3) 多进程技术

Performer 提供了一系列的多进程技术,将视觉仿真中四个阶段 App, Cull, Draw 以及 Intersect 分配不同的进程处理,如表 41-4-5 所示。

通过调用 pfMultiProcess() 函数分配 App, cull, draw 阶段,需要注意进程分配最好结合到处理器的分配。

表 41-4-5 多进程模型

进程名称	说明
PFMP_APPCULLDRAW	绑定 app, cull, draw 在一个进程中
PFMP_APP_CULLDRAW	绑定 cull, draw 在一个进程, app 在一个进程中
PFMP_FORK_CULL	绑定 app, cull 在一个进程, draw 在一个进程中
PFMP_APP_CULL_DRAW	绑定 app, cull, draw 分配在不同的进程中
PFMP_FORK_CULL_DRAW	
PFMP_FORK_DRAW	

#### 2.1.7 可视化效果

Performer 提供了强大的视觉处理功能,输出了环境、大气、天空等可视化效果增加场景真实性。

大地、天空以及地平线:使用 pfEarthSky 函数。大气层效果:配置 pfEarthSky 的一些属性,可以做到非常复杂的大气层效果,如图 41-4-15a。

各层代表的效果分别为地表雾气 (ground fog), 天空和云气 (clouds), 最低层为雾气,范围可以由用户指定,高度增加雾气逐渐减弱,如图 41-4-15b。默认情况下,大地、天空模型不绘制,通道只是简单清屏为黑色,将 Z-buffer 设置为最大值。

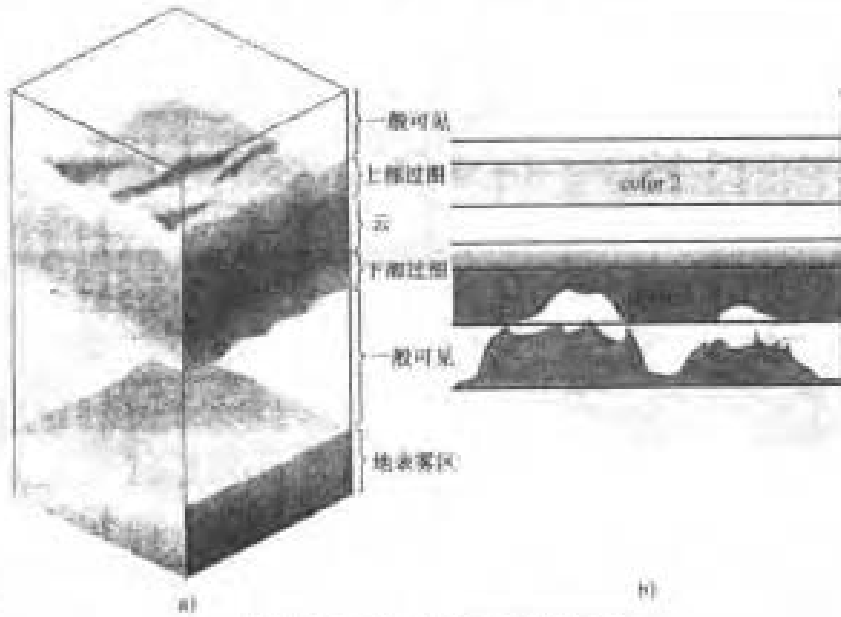


图 41-4-15 多层大气和层次化场景模型

## 2.2 基于 Performer 的实时仿真系统开发框架

### 2.2.1 场景创建的基本框架

一个 OpenGL Performer 程序只需几十行代码，而且其基本结构与复杂程序基本相同，因此一旦掌握

应用程序开发的基本框架，就可以进一步开发更为复杂的 OpenGL Performer 程序。

基本框架都是从配置 channel、pipe 定义窗口，然后自动仿真循环，如图 41-4-16 所示。Performer 提供两个应用开发框架，一是基于 libpf 的应用框架，一是基于 libpfv 的开发框架，这是 OpenGL Performer 3.0 以后新提供的，具有更大的方便性。

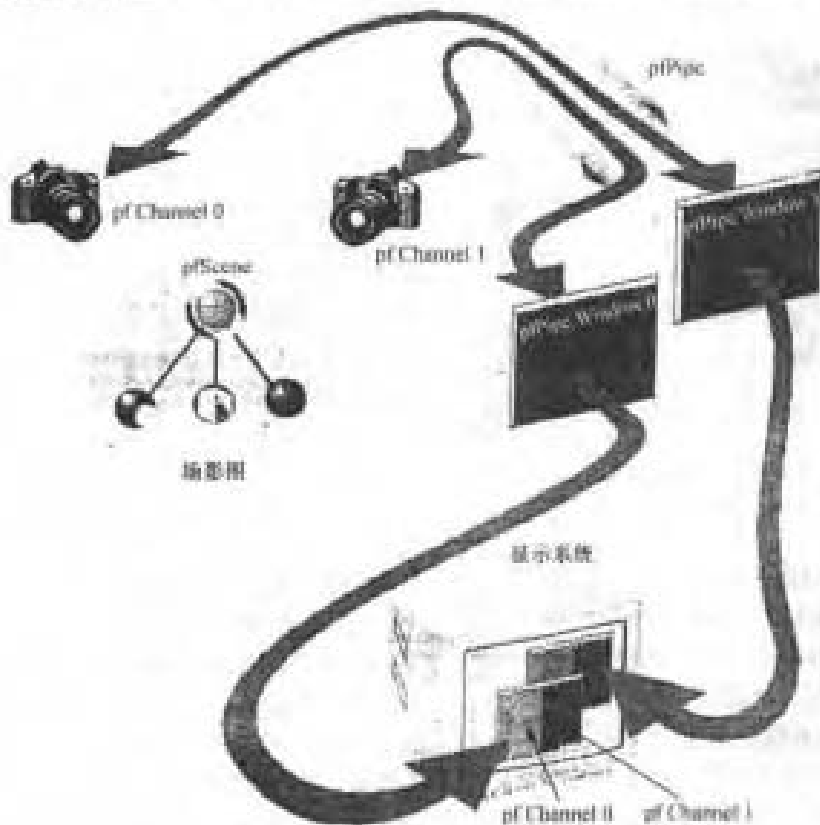


图 41-4-16 Performer 应用开发元数据流程

## 2.2.2 基于 libpf 应用程序的结构分析

使用 libpf 库, 利用数十行代码就可以构建一个标准的 Performer 应用程序。

## (1) 简介

## (2) 应用程序代码及其注解

```
#include <stdlib.h>
#include <Performer/pf.h>
#include <Performer/pfutil.h>
#include <Performer/pfdu.h>

int
main (int argc, char *argv[])
{
    float t=0.0f;
    pfScene *scene;
    pfNode *root;
    pfPipe *p;
    pfPipeWindow *pw;
    pfChannel *chan;
    pfSphere bsphere;

    if (argc<2)
    {
        pfNotify (PFNFY_FATAL, PFNFY_USAGE, "Usage: simple file.ext\n");
        exit(1);
    }

    /* 初始化 Performer */
    pfInit();

    /*
     * 根据处理器数量选择使用默认的多进程模式
     */
    pfMultiprocess (PFMP_DEFAULT);

    /* 在 pfConfig()进程启用前加载所有 DSO 载入器 */
    pfdInitConverter(argv[1]);

    /*
     * 初始化 pfMultiprocess 中调用的多进程模式派生 Performer 的进程
     * CUI.L、DRAW, 等
     */
    pfConfig();

    /*
     * 设置 Performer 搜索路径
     */
    pfFilePath(".", "/usr/share/Performer/data");

    /* 读入一个已知类型的文件 */
    if ((root = pfdLoadFile(argv[1])) == NULL)
```



```
{
pfExit();
exit( 1)

/* 将载人的文件附着在一个视景上 */
scene = pfNewScene();
pfAddChild(scene, root);

/* 在视景上, 建立光源 */
pfAddChild(scene, pfNewLSource());

/* 设置管道, 建立管道和窗口间联系 */
p = pfGetPipe(0);
pw = pfNewPWin(p);
pfPWinType(pw, PFPWIN_TYPE_X);
pfPWinName(pw, "OpenGL Performer");
pfPWinOriginSize(pw, 0, 0, 500, 500);

/* 设置并建立 GL 窗口 */
pfOpenPWin(pw);

/* 建立配置管道 */
chan = pfNewChan(p);
pfChanScene(chan, scene);
pfChanFOV(chan, 45.0f, 0.0f);

/* 设置视锥 */
pfGetNodeBSphere (root, &bsphere);
pfChanNearFar(chan, 1.0f, 10.0f * bsphere.radius);

/* 设置 20 秒仿真循环 */
while (t < 20.0f)
{
    pfCoord view;
    float s, c;

    /* 计算新位置 */
    t = pfGetTime();
    pfSinCos (45.0f * t, &s, &c);
    pfSetVec3(view.hpr, 45.0f * t, -10.0f, 0);
    pfSetVec3(view.xyz, 2.0f * bsphere.radius * s,
        -2.0f * bsphere.radius * c,
        0.5f * bsphere.radius);
    pfChanView(chan, view.xyz, view.hpr);
    /* 为本帧初始化 cull/draw 进程 */
    pfFrame();
}

/* 终止并行进程, 退出 */
pfExit();
}
```

## (3) 结构解析

## 1) 必要的头文件。

```
#include <stdlib.h>
```

## 2) 相关 OpenGL Performer 头文件。

```
#include <Performer/pf.h>
#include <Performer/pfutil.h>
#include <Performer/pfdu.h>
```

## 3) 声明所需变量。

pfScene	定义一个在通道上被渲染的视景图形
pfPipe	定义一条执行渲染操作的流水线 Pipeline(管道)
pfChannel	定义一个在指定流水线上被渲染的视图 Channel(通道)

4) 可以为每个管道配置多个通道,多个窗口用来定义多个视景。

## 5) 初始化 OpenGL Performer。

```
pfInit();
```

设置供多进程使用的共享内存区,初始化高精度时钟,并置 OpenGL Performer 的初始状态。

## 6) 配置 OpenGL Performer。

```
pfConfig();
```

配置流水线条数,并根据所选择的多进程模式开始处理。

## 7) 加载或创建可视化数据库。

```
root = pfdLoadFile(argv[1]);
```

根据给定文件的扩展名自动选择合适的载入器。

## 8) 创建新的视景。

```
scene = pfNewScene();
```

9) 将可视化数据库的根节点加入视景。

```
pfAddChild(scene, root);
```

## 10) 初始化图形流水线并配置窗口。

```
p = pfGetPipe(0);
pw = pfNewPWin(p);
pfPWinType(pw, PFPWIN_TYPE_X);
pfPWinName(pw, "OpenGL Performer");
pfPWinOriginSize(pw, 0, 0, 500, 500);

/* 建立回调函数 OpenPipeWin() 的配置 */
pfPWinConfigFunc(pw, OpenPipeWin);

/* 打开和配置窗口 */
pfConfigPWin(pw);
```

## 11) 指定帧率和同步模式。

```
pfFrameRate(60);
pfPhase(PFPHASE_LOCK);
```

## 12) 在指定管道上创建通道。

```
chan = pfNewChan(p);
```

一条通道是一条管道中的一个视口。

## 13) 配置通道。

```
pfChanScene(chan, scene);
pfChanFOV(chan, 45.0f, 0.0f);
pfGetNodeBSphere(root, &bsphere);
pfChanNearFar(chan, 1.0f, 10.0f * bsphere.radius);
```

设置视点,视野(FOV)和远近剪裁平面。

## 14) 重复渲染视景直至仿真时间结束。

为下一帧设置视点:

```
pfChanView(chan, view.xyz, view.hpr);
```

为渲染下一帧初始化渲染 CULL/DRAW 进程:

```
pfFrame();
```

## 15) 仿真时间结束,退出。

```
pfExit();
```

## 2.2.3 基于 libpfv 应用程序的结构分析

## (1) 简介

Libpfv 库中最主要的类是 pfvViewer。它允许复杂的多场景和多视口的应用,这些应用可以模块化,按照面向对象原则封装,极大满足程序的配置和重用。

对于 libpfv,OpenGL Performer 提供以下特性:

- 加载几何物体到 pfvViewer 场景中。
- 通过鼠标拾取场景中几何物体。
- 操纵几何物体(缩放,移动,旋转,删除)。
- 通过鼠标和键盘漫游场景。
- 控制模型的渲染模式。
- 自由设置地面、天空多种背景。
- 覆盖方式显示 2D 图像。
- 可以保存实时渲染的快照。
- 视景的切换平滑。
- 搜集和显示系统统计数据。

## (2) 程序结构分析

pfvViewer 比 libpf 提供更为简单的编程接口。

## 1) 通过 API 调用建立场景源代码:

```

#include <Performer/pfdu.h>
#include <Performer/pf/pfLightSource.h>
#include <Performer/pfv/pfvViewer.h>

main (int argc, char *argv[])
{
    //初始化 Performer
    pfInit();

    //实例化一个 pfvViewer
    pfvViewer * viewer=new pfvViewer;

    //初始化调入模型文件
    pfdInitConverter(argv[1]);

    //配置、初始化 pfvViewer
    viewer->config();

    //场景中加入光源
    viewer->addNode (new pfLightSource);

    //加载模型到场景中
    viewer->addNode (pfdLoadFile(argv[1]));

    //启动仿真循环
    viewer->run();
}

```

2) 通过 XML 配置文件建立场景源代码;  
这种方法应该是 Performer 划时代的改变,通过

XML 配置将应用扩展开来,程序也变得非常可读。

```

#include <Performer/pfv/pfvViewer.h>

Int main (int argc, char *argv[])
{
    //初始化 Performer
    pfInit();

    //实例化 pfvViewer 并建立读入 XML 配置文件接口 argv[1]
    pfvViewer * viewer=new pfvViewer(argv[1]);

    //配置并初始化 pfvViewer.
    viewer->config();

    //启动仿真循环
    viewer->run();
}

```

最简单的 XML 配置文件格式如下：

```
<? xml version="1.0"?>
<viewer>
  <module>
    <! --Loader module; loads models into world -->
    <class>pfvmLoader</class>
    <data>
      <Model>
        <FileName>esprit. flt</FileName>
      </Model>
    </data>
  </module>

  <! --Add a trackball navigation module -->
  <module>
    <class>pfvmTrackball</class>
  </module>

  <! -- Add a picking module -->
  <module>
    <class>pfvmPicker</class>
  </module>
</viewer>
```

更为复杂 XML 配置文件,包括显示标签用来建立复杂的显示配置下列例子用来设置三个通道的全景视图:

```
<? xml version="1.0"?>
<viewer>
  <! --Display specifications -->
  <display>
    <! --Configure middle pipe -->
    <pipe>
      <! --direct middle pipe to screen 0 -->
      <screen>0</screen>
      <! -- Configure a single pipe-window on middle pipe - >
      <pwin>
        <! - set pipe-window to fullscreen, no border -->
        <fullscreen>1</fullscreen>
        <border>0</border>
        <! -- Configure a single channel on pipe-window -->
        <chan>
          <viewrange>0.32, 0.68, 0.0, 1.0</viewrange>
          <hprOffset>0.0, 0.0, 0.0</hprOffset>
          <fov>59.0, 46.0</fov>
        </chan>
      </pwin>
    </pipe>

    <! --Configure right pipe -->
    <pipe>
      <screen>1</screen>
      <pwin>
        <! -- set pipe-window to fullscreen, no border -->
```

```

    <fullscreen>1</fullscreen>
    <border>0</border>
    <! -- Configure a single channel on pipe-window -->
    <chan>
        <viewrange>0, 64, 1, 0, 0, 0, 1, 0</viewrange>
        <hprOffset> 53, 333, 0, 0, 0, 0</hprOffset>
        <fov>59, 0, 46, 0</fov>
    </chan>
</pwin>
</pipe>

<! Configure left pipe \
(pipe)
    <screen>2</screen>
    <pwin>
        <! -- set pipe-window to fullscreen, no border -->
        <fullscreen>1</fullscreen>
        <border>0</border>
        <! - Configure a single channel on pipe-window -->
        <chan>
            <viewrange>0, 0, 0, 36, 0, 0, 1, 0</viewrange>
            <hprOffset>53, 333, 0, 0, 0, 0</hprOffset>
            <fov>59, 0, 46, 0</fov>
        </chan>
    </pwin>
</pipe>
</display>

<module>
    <! - Loader module; loads models into world -->
    <class>pfvmLoader</class>
    <data>
        <Model>
            <FileName>esprit.flt</FileName>
        </Model>
    </data>
</module>

<! - Add a trackball navigation module -->
<module>
    <class>pfvmTrackball</class>
</module>

<! Add a picking module -->
<module>
    <class>pfvmPicker</class>
</module>

</viewer>

```

图 41.4-17 是一个基于 libpfv 的应用开发例子, 可以看到轿车模型的轮胎可以被拾取和移动。



# 第5章 工程分析数据沉浸可视化技术

## 1 工程分析数据沉浸可视化技术概述

### 1.1 科学计算可视化的概念

科学计算可视化 (Visualization in Scientific Computing), 指的是运用计算机图形学和图像处理技术, 将科学计算过程中及计算结束后产生的数据转换为图形或图像在屏幕上显示出来并进行交互处理的理论方法技术。该概念于1987年由美国国家基金会第一次提出。随着科学技术的发展, 由超级计算机、测量仪器、人造地球卫星、航天飞机、地震勘探等产生的数据量越来越大。存放数据的磁带堆满了幢幢大楼, 只有将它们转换为图像才能理解其包含的信息、了解其规律、甚至发现用其他方法不能发现的现象。否则, 只能将这些数据收集起来并束之高阁。

科学计算可视化是计算机图形学的一个重要领域, 它的核心是将三维数据转换为图像, 它涉及到标量、矢量、张量的可视化、流场的可视化、数值模拟及计算的交互控制、海量数据的存储、处理及传输、图形及图像处理的向量及并行算法等。

目前应用广泛, 如医学及医疗、地震勘探、气象预报、分子结构、流体力学、有限元分析、天体物理、海洋观察、地理信息、洪水预报、环境保护等。

随着虚拟样机技术的发展, 如何应用科学计算可视化技术来表达虚拟样机的功能、性能, 显得尤为重要。虚拟现实技术的融合更为科学计算可视化带来了新的应用领域。所谓工程分析数据沉浸可视化指融入虚拟现实技术的工程分析数据可视化。利用这种沉浸可视化的方法来显示工程分析数据, 有助于使设计人员更直观、更准确地理解数据, 更快速容易地找出设计中存在的问题, 从而也加快了产品设计开发的周期。

工程分析数据可视化系统的意义在于以下因素:

1) 诸如 MSC/Nastran 等工程分析软件虽然提供了对分析结果进行后处理的可视化模块, 但是可视化功能仍有限, 使设计人员难以发现大型、复杂模型中的细小、隐蔽问题。

2) 一个产品的开发往往是需要考察多种性能, 各种不同的性能需要不同的分析工具实现, 将多种分析数据集成在一起, 使得人们认识产品的性能是站在一个全局的视图上, 而不是局部。

3) 虚拟现实 (VR) 技术的引入, 使得人们认识产

品性能提高了一个层次, 现有分析软件都不能接入 VR。

### 1.2 科学计算可视化的研究内容

科学计算可视化可按功能分为三个层次:

#### (1) 科学计算结果数据的后处理

将计算过程和可视化过程分开, 在脱机状态下对计算的结果数据或测量数据实现可视化。由于不要求实时地用图形、图像显示数据, 因而这一层次的可视化的功能对计算功能的需求较之下面两个层次要低一些。

#### (2) 科学计算结果数据的实时处理和显示

在进行科学计算的同时, 实时地对计算的结果或测量数据实现可视化。这一层次的功能较之上一个层次需要更强的计算能力。

#### (3) 科学计算结果数据的实时绘制和交互处理

在这里, 绘制 (rendering) 指的是由物体的几何模型生成屏幕图像的过程。这一层次的功能不仅能对数据进行实时处理和显示, 而且如有必要, 还可以通过交互方式修改原始数据、边界条件或其他参数, 使计算结果更为满意, 实现用户对科学计算过程的交互控制和引导。很显然, 这一层次的功能不仅要求计算机硬件具有很强的计算能力, 而且要求可视化软件具有很强的交互功能。

### 1.3 科学计算可视化技术发展趋势

科学计算可视化将图形和图像技术应用于科学计算, 不仅需要分析由计算机得出的计算数据, 而且需要了解在计算过程中数据的变化, 而这些都需要借助于计算机图形学及图像处理技术。由于科学计算可视化可以将计算结果用图形或图像形象、直观的显示出来, 因而许多抽象的、难于理解的原理和规律变得容易理解了 (例如, 不稳定流场可以用迹线直观地表示出来), 许多冗繁而枯燥的数据变得生动有趣了。

科学计算可视化运用计算机图形学和图像处理技术, 将数据转换为图形或图像在屏幕上显示出来, 并进行交互处理的理论、方法和技术。它涉及到计算机图形学、图像处理、计算机辅助设计、计算机视觉及人机交互技术等多个领域。科学家们不仅需要通过图形图像来分析由计算机算出的数据, 而且需要了解在计算过程中数据的变化。

科学计算可视化技术本身目前正朝着超级计算、海量数据处理和虚拟现实技术的融合以及在工程领域中的进一步应用快速发展。

## 2 数据模型准备

工程设计过程中,大量应用到三维 CAD 模型,典型的 CAD 面模型是通过数学的方法进行描述的,通常都会利用非均匀有理 B 样条 (Non-Uniform Rational B-Splines, NURBS) 样条曲线,这种描述方法有利于对设计曲面进行操纵与修改。NURBS 曲面具有极好的连续性,多边形网格曲面虽然也能保持一次连续性,但面与面之间存在棱边,利用 NURBS 的方法能将数据存储在更小的内存空间和更小的文件中,通常一个完整的模型不是由一个 NURBS 曲面构成的,而是通过多个 NURBS 曲面连接而成,这样每一个 NURBS 曲面被称为 NURBS 块。这些模型一方面用来几何外形、结构设计,一方面用于分析 (CAE),而虚

拟现实环境中的模型是基于面模型,并非是 NURBS 曲面,需要将 CAD 的表面或实体模型数据格式转换成多边形面片,以满足虚拟环境中场景的实时绘制需求。这种转换通常会导致在计算出的网格中出现边界处面片的不连续、错误的面法矢和一些洞孔,基于多边形的数据格式转换在工程虚拟现实应用是一件非常重要的工作。

目前大多 CAE 分析软件工具的数据(标量或矢量数据)可视化方式还只限于三维数据的两维投影,不是特别直观,因此也有需要在虚拟环境下进行 CAE 数据的沉浸可视化。因而也同样存在 CAE 数据到虚拟环境的转换。

如图 41.5-1 所示,通过 CAD 软件造型后的三维数据需要进行数据格式的转换和数据量的精简;经 CAE 计算后得到的标量或矢量数据也需经过适当的转换。转换成满足虚拟环境需求的一些较为通用的数据格式,常见的如 VRML 等。

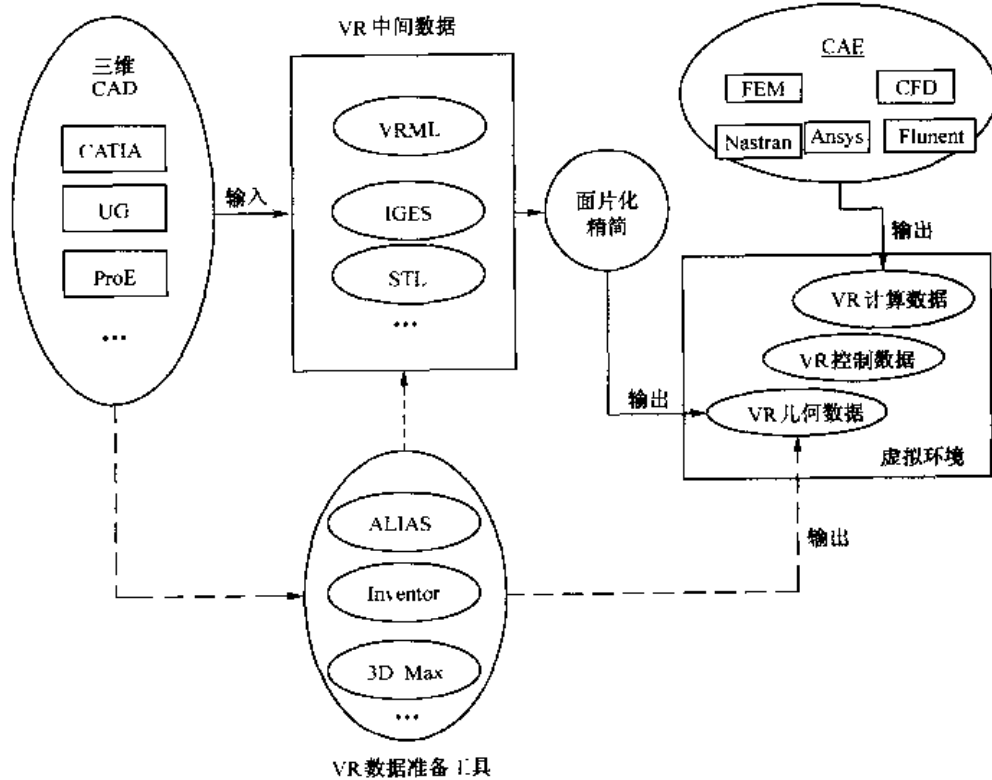


图 41.5-1 到虚拟环境的数据转换与准备

### 2.1 CAD 数据的精简

将基于 NURBS 的 CAD 数据转换成多边形的数据,是一个网格化的过程,为了能更逼真地表现出模型,通常将产生成千上万的多边形。对于现今的计算机硬件水平,有必要将多边形的数量控制在 10 000~80 000 个之内。这种精简是非常必要的,在 VR 环境

中,如果需要产生一个具有沉浸感的场景必须保持每秒 10~25 帧的计算速度。

对于 CAD 数据,存在两种不同的渠道来限制多边形的数量。数学上的任意形式的曲面不存在一个统一的多边形描述的方法。因此,一个可行的方法是在网格划分的时候,选择一个适合的网格数量,从而使得到的多边形不会超过某一特定的数量值。大多数的 CAD 系统也允许定



义基于多边形的数据格式。但是如果没有一个多边形的精简算法就无法对这种数据进行删减。而且,多数情况下,这种基于多边形的数据的精简算法也能应用于其他众多的领域。通过高质量的网格剖分,先得到一个精确的

多边形面片的模型,然后将其删减以控制其复杂程度。如图 41-5-2 所示,对驾驶员座椅进行多边形精简,多边形数由原来的 3517 缩减为 1957,外观并无太大的变化,但更能满足 VR 实时性的需求。

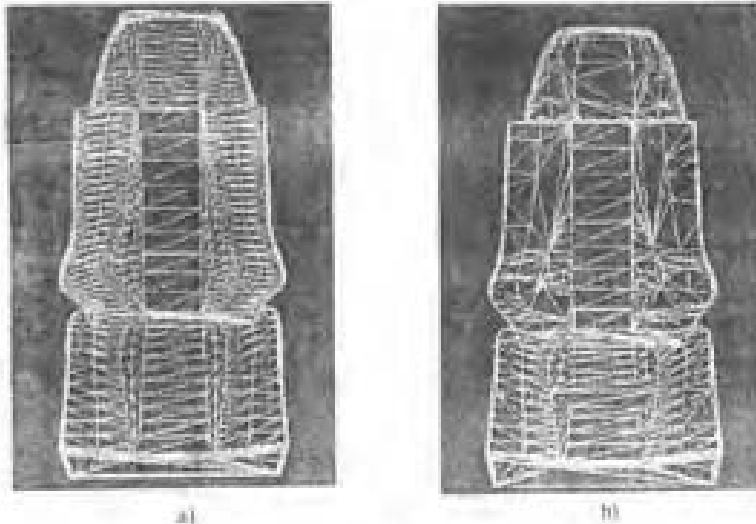


图 41-5-2 驾驶员座椅进行多边形删减前后的对比

a) 简化前 b) 简化后

目前已有不少常用商品化软件具有数据删减的功能,但对一些特殊定义格式的 CAD 数据,还是需要特定的数据精简工具来完成数据的简化。本节只讨论边界折叠法的实现,这种方法通过反复应用简单的边界折叠操作来达到降低模型复杂度的目的。如图 41-5-3 所示。在这种操作中,两个结点  $u$  和  $v$  (边界  $uv$ ) 被选中,其中一个  $u$  被“移动”或“折叠”到另一个  $v$  (在本例中设为  $v$ )。

操作步骤为:

- 1) 删去所有既包括  $u$  又包括  $v$  的结点 (也就是删去边界  $uv$  上的三角形)。
- 2) 将所有包含结点  $u$  的剩下的多边形更新为用结点  $v$  替代  $u$ 。

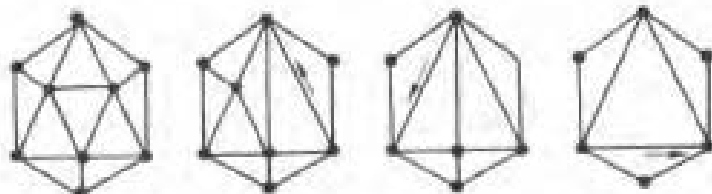


图 41-5-3 最简单的边界折叠操作

a) 操作前 b) 操作后

图 41-5-4 边界折叠操作的例子 (从左到右共三次折叠)

生成好的低多边形模型关键是选择的折叠边界应尽可能少地改变模型的外观。研究者提出了各种各样的方法来决定每一步所应折叠的“最小代价”的边界。一般最好的方法非常复杂(很难实现)并且要花费太长的时间计算。下面介绍一种简单而快速的选择方法,可以得到相当好的低多边形模型。

很明显,首先去掉小细节是可行的。注意到,接近共面的表面需要较少的多边形来表达,而高曲率的表面需要较多的多边形。故用边界长度乘以曲率项来定义折叠边界的代价。折叠边界  $uv$  的曲率项是通过比较表面法向量的点积找出与  $u$  相邻而离  $uv$  的其他三角形最远的点。式 (41-5-1) 以更正规的形式显示了

边界代价公式:

$$\text{cost}(u,v) = |u - v| \times \max\{\min((1 - f_{\text{normal}} + a_{\text{normal}}) \div 2), f_{\text{area}}\} \quad (41.5-1)$$

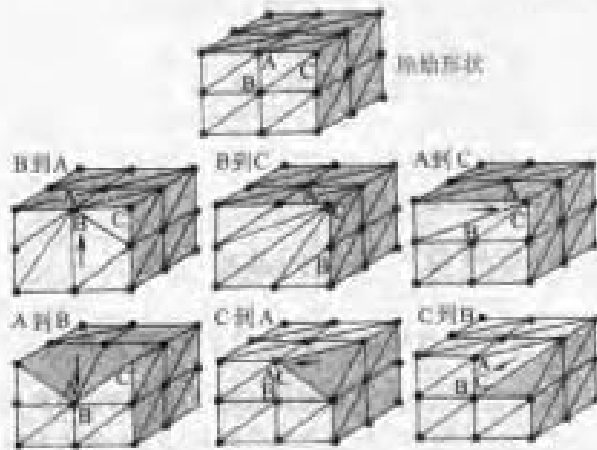


图 41.5-5 不同缩叠方法的比较

从式(41.5-1)中可以看出这一算法在决定缩叠边界时兼顾了曲率和大小。请注意,将结点  $u$  缩叠到结点  $v$  与将结点  $v$  缩叠到结点  $u$  的代价是不同的。此外,这个公式对于沿棱的缩叠是有效的。图 41.5-5 说明了这个问题。很明显,在平面中点的结点  $B$  可以缩叠到  $A$  或  $C$ 。顶点  $C$  应该保留,如果把上棱边的结点  $A$  缩叠到内部结点  $B$  效果将很差。但是,  $A$  可以沿棱边移到  $C$  上而不影响整个模型形状。

下面是决定边界缩叠的代价和边界缩叠操作的三个函数描述。

```
float ComputeEdgeCollapseCost(Vertex * u, Vertex * v) {
```

(计算通过把  $u$  移到  $v$  来缩叠边界,那么模型将改变多少)

- 找出在边界  $uv$  上的三角形
- 用离河边最远的三角形面来决定曲面项
- 用面法向量的点积
- 返回边界长度

```
void ComputeEdgeCostAtVertex(Vertex * v) {
```

寻找所有“最少代价”边的邻边

```
void Collapse(Vertex * u, Vertex * v) {
```

- 通过将结点  $u$  移到  $v$  来缩叠边界  $uv$
- 若  $u$  是一个单独的结点,做将它删去
- 组织  $u$  的所有相邻列表
- 删除  $uv$  边上的三角形

将剩下的三角形更新为以  $v$  代替  $u$   
重新计算邻边的边界缩叠代价

这一算法只能对三角形进行处理,三边以上的多边形可以先三角形化,再用该算法简化。有了这些函数多边形缩减就容易实现了,只要用物体的几何形状初始化结点和三角形列表,然后执行类似下面的语句:

```
Vertex * mn = MinimumCostEdge();
Collapse(mn, mn->collapse);
```

## 2.2 CAE 数据的转换

工程分析中的 CAE 计算仿真分析如 FEA、CFD 等产生的仿真数据,种类繁多,不同的软件系统会有不同的数据定义方式,如表 41.5-1 和表 41.5-2 所示。

表 41.5-1 有限元分析 (FEA) 软件产生的数据格式

软件工具	数据格式
ABAQUS	Direct 格式 for binary 或者 ascii .in 文件
ANSYS	Direct 格式 for binary: .rst, .rth, .rta, .rtd 文件
FENSAP	可以输出多种文件,用户自定义读取
I-DEAS	Translator for I-DEAS Universal file
LS-DYNA	用户自定义格式 for d3plot 文件
MSC/Dytran	用户自定义格式 for MSC/Dytran Archive (.arc) 文件
MSC/Nastran	用户自定义格式 for binary OP2 文件
PATRAN	Converts PATRAN neutral 文件 to MOVIE, BYE 格式
ADAGIO	可以输出多种文件,用户自定义读取
ALEGRA	可以输出多种文件,用户自定义读取

表 41.5-2 计算流体力学分析 (CFD) 软件产生的数据格式

软件工具	数据格式
CFD++	Exports EnSight Case 格式
CFD-ACE	DTF 格式
CFD-FASTRAN	DTF 格式
CFDESIGN	Terplot 格式
FAST Unstructured	NASA FAST unstructured 格式
FIDAP	FIDAP neutral (FDNEUT) 文件
FINE/Aero	PLOT3D 或者 CGNS 文件/格式
FINE/Turbo	PLOT3D 或者 CGNS 文件/格式
FIRE	Code EnSight 格式

(续)

软件工具	数据格式
Flow-3D	用户自定义格式 f 或者 FLOW 3D results (flsgrf) 文件
FLUENT	Converts Fluent Particle file to EnSight 格式
Fluent	Case 或者 EnSight Casefile
POAM	Contact developer (Imperial College)
PHOENICS	PLOT3D 文件/格式/ensight
PLOT3D	Direct 格式 f 或者 PLOT3D and FAST structured f 或者 mats
STAR-CD	Code EnSight Casefile

由于现有的科学数据的种类繁多,数量大,因而把所有的数据都转换为一种标准格式是不现实的。然而,为方便在不同的应用系统之间进行数据交换和共享,仍然有必要建立标准的科学数据标准格式。目前,研究人员已提出多种数据格式,其中常用的有层次数据格式 HDF (Hierarchical Data Format),公共数据格式 CDF (Common Data Format),还有网络公共数据格式 NetCDF 等,如表 41-5-3 所示。HDF、NetCDF 等数据格式是一种通用的数据格式,通用性在带来广泛的兼容性的同时,必然在用于特定目的时(如仅矢量场可视化)产生效率低下、或至少不是最优的问题。

表 41-5-3 一些通用数据格式

数据格式	描述
ENSIGHT5	Original, unstructured EnSight format
CASE(ENSIGHT6)	Native EnSight format
HDF	NASA 层次数据格式标准
MOVIE.BYU	MOVIE.BYU 格式
NetCDF	NASA 通用网络数据传输标准
POLY-3D	变形数据格式
SILO	SILO API 通用读取数据标准
STL	STL 几何数据格式
TECPLOT	TECPLOT 结构和非结构数据标准
Telluride	EnSight 格式
VTK	Visualization toolkits 数据格式
Plot3D	NASA 数据格式 (二进制/ASCII 格式)
XML	扩展标志语言,通用的数据交换标准
VRML/X3D	网络图形交互标准

FEA 软件 (Ansys, Nastran, Patran, Dytran 等) 以及一些计算流体动力学软件 (StarCD, Fluent, CFX) 等产生的数据有矢量场数据也有标量场数据,这些数据往往满足某些特定要求,如果不进行二次处理,比较难满足虚拟环境中的可视化要求。通过仔细分析上述数据的结构类型,在通用的数据标准 VRML、Plot3D、NetCDF、VTK 以及 XML 的基础上,构建了面向矢量场和标量场通用的可以扩展的数据交互标准,用于后续章节的矢量场数据可视化。用于科学计算

可视化的数据应能满足下述要求 (部分或全部):

- 1) 提供读/写数据的一般方法。
- 2) 对数据进行有效组织,以方便数据的处理。
- 3) 用户可接受的响应时间,用户能与数据交互。
- 4) 数据管理和单纯的数据访问分开。
- 5) 冗余性控制。
- 6) 能更改传递,避免不一致。
- 7) 存档设施。
- 8) 安全性策略。
- 9) 数据完整性的维护。

### 3 科学计算可视化的基础技术

#### 3.1 可视化数据的组织形式及物理分类

科学计算可视化的数据由有限空间内的离散采样数据点构成,每个采样点上的采样值代表在该点上的一个或多个物理属性值。可视化处理的数据,第一类是来自对现实世界的测量结果,如医学上的计算机断层扫描 (CT) 数据和核磁共振 (MRI) 数据;第二类来自对现实世界进行模拟计算产生的数据,如工程上有限元分析和计算流体动力学的方法产生各种模拟数据;第三类来自对现实世界相关的物理量进行计算处理后产生数据,这些数据并不是实际存在物理量,如工程中常用变换域计算的结果,如信号处理中的离散傅里叶变换 (DFT)、离散余弦变换 (DCT)、离散小波变换 (DWT) 等。

根据数据的类型进行划分,数据场可以分为标量场和矢量场。标量场是指数据只有大小而无方向的数据场。比较常见的有密度场、温度场等,标量场的可视化主要是揭示各分类物质的空间分布;矢量场(也称向量场)数据不仅有数值的大小,还有方向的变化,典型矢量场是流场。矢量场的可视化除了揭示各类物质的空间分布外,还要反映其变化的趋势。本节重点关注三维矢量场的可视化,并以来自于计算流体动力学 (CFD) 的数值仿真作为例子来说明。

矢量场可视化的基础是如何描述矢量的数据类型。随着应用领域的不同,数据来源的不同,所需要观察的数据也很不相同。而实现三维空间数据场可视化的算法却与数据类型有极大的关系。数据类型有两层含义,一是数据本身的类型,二是数据分布及连接关系的类型。

##### (1) 数据本身的类型

数据的类型可以分成三类:只有大小没有方向的标量、既有大小又有方向的矢量和具有多重方向性的张量。

- 1) 标量 标量只有一个分量,只有大小没有方

向,因此它与坐标转换没有相关性:设标量  $x$  在新、老坐标系的值是  $x'$  和  $x$ , 则

$$x' = x \quad (41.5-2)$$

常见的标量有时间、温度、密度。

2) 矢量 常见的矢量有位移、速度、加速度等。

矢量有两个分量,一个是表征大小的数字,另一个是其方向。矢量需要根据参考坐标系中分解来描述,但其性质与坐标系无关。例如图 41.5-6 中的矢量  $X$  分解式为:

$$X = x_1 e_1 + x_2 e_2 + x_3 e_3 \quad (41.5-3)$$

式中  $x_1, x_2, x_3$  为 矢量  $X$  的三个分量;

$e_1, e_2, e_3$  一沿坐标轴的三个单位基矢量。

因此矢量  $X$  可以用分量记法表示为  $x_1, x_2, x_3$ 。

$x_1, x_2, x_3$  与  $x_1', x_2', x_3'$  分别是  $X$  在旧坐标系和新坐标系中的投影,显然  $x_1, x_2, x_3$  与  $x_1', x_2', x_3'$  之间有以下关系:

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned} \quad (41.5-4)$$

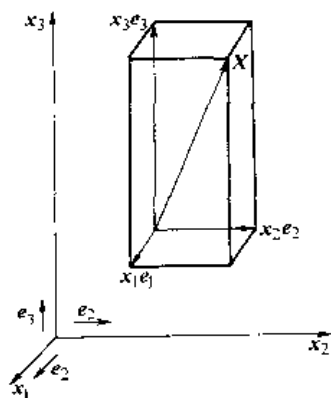


图 41.5-6 矢量  $X$  的表达

可根据爱因斯坦求和约定简写为:

$$x_i' = a_{ij}x_j \quad (41.5-5)$$

3) 张量 将矢量按以坐标变换为基础的定义加以推广,即可得到张量定义.任意二阶张量的分解式可以表示为:

$$\begin{aligned} X &= x_{11}e_1e_1 + x_{12}e_1e_2 + x_{13}e_1e_3 + \\ & x_{21}e_2e_1 + x_{22}e_2e_2 + x_{23}e_2e_3 + \\ & x_{31}e_3e_1 + x_{32}e_3e_2 + x_{33}e_3e_3 \\ & = x_{ij}e_i e_j \end{aligned} \quad (41.5-6)$$

式中  $x_{ij}$  一 张量分量;

$e_i, e_j$  一 基张量,它是两个基矢量,写在一起而不做任何运算,构成张量的基。

二阶张量的定义可以推广到  $n$  阶张量中去。当  $n=0$  时,张量的分量只有一个,它是一个标量。因此,

可将标量视为零阶张量。 $n=1$ ,是一个矢量,矢量可视为一阶张量。张量可用于流体力学的计算中,用以表示流体微团的微观变化。

(2) 数据分布及连接关系的类型

在科学计算中,所研究对象的特性往往是用一组方程式来描述的,通常为常微分方程、偏微分方程、线性或非线性方程。如果能求出这些方程组的解析解,那么就可以在这些方程组所定义的空间内的任意位置得到所需要的解。但是,只有当方程组非常简单时,才具有这种可能性。一般情况下,只能得到数值解。为此需要所定义的空间离散化,离散成体单元、面单元、线段或者网格点,再用数值求解方法,求出这些离散单元的函数值。所以科学计算的结果数据多是离散的,而不是连续的。研究的可视化对象指的是空间上离散的三维矢量数据。三维空间上离散数据之间的连接关系主要有以下类型:

1) 结构化数据 在逻辑上组织成三维数组的空间离散数据,每个元素都可以有它自己所在的层号、行号和列号。根据结构化数据中各元素的物理分布,又分为:

均匀网格化数据——数据均匀分布在三维网格点上,无需给出各个点的空间位置,只要给出其中一个点空间位置和数据点序号,就可以根据网格间距求出其他点,如图 41.5-7a。

规则网格结构化数据——数据分布在长方体的三维网格上,长方体的大小各不相同,无规则可循,其可视化除了给出三维数组的下标,还要给出它所在的空间坐标系,如图 41.5-7b。

不规则网格结构化数据——数据在逻辑上被组织成三维数组,但是在空间位置的分布没有任何规则,这类网格数据,需要知道每个点的三维数组下标,还要给出所在的空间坐标,有限元计算经常就是这种类型数据,如图 41.5-7c。

2) 非结构化网格 由一系列的单元组成,但是不能组成三维数组。这些单元由四面体、六面体、三棱柱或者四棱锥构成,单元的面可以为平面也可以是曲面,边可以是直线段也可以是曲线段。数据类型常出现在计算流体动力学中。必须给出每个数据点的空间位置,同时还有其互相间的拓扑关系,如图 41.5-7d。

3) 混合性数据 结构化数据和非结构化数据混合在一起,以处理某些特殊类型,如图 41.5-7e。

### 3.2 可视化流程

可视化的流程一般分为四个步骤:数据生成、数据精炼及过滤、数据映射和绘制。其可视化流程如图 41.5-8 所示。

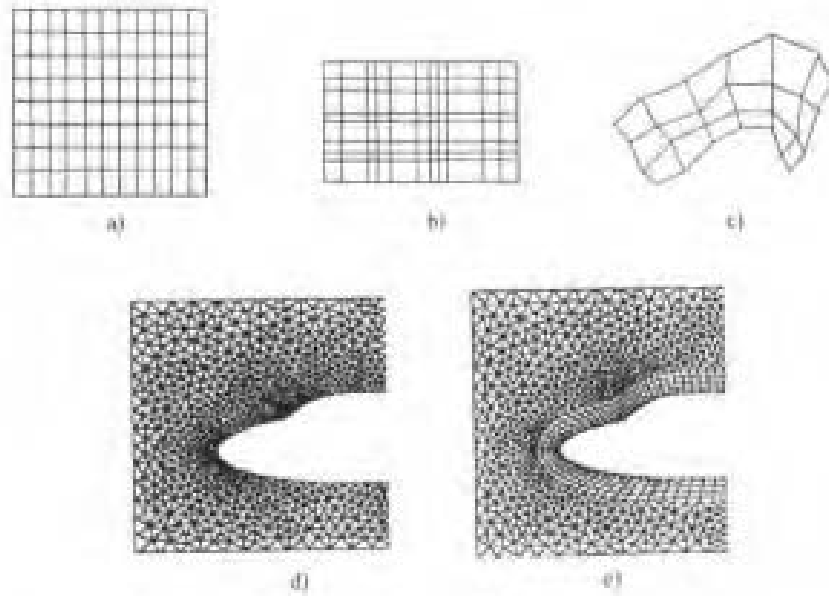


图 41.5.7 网络类型

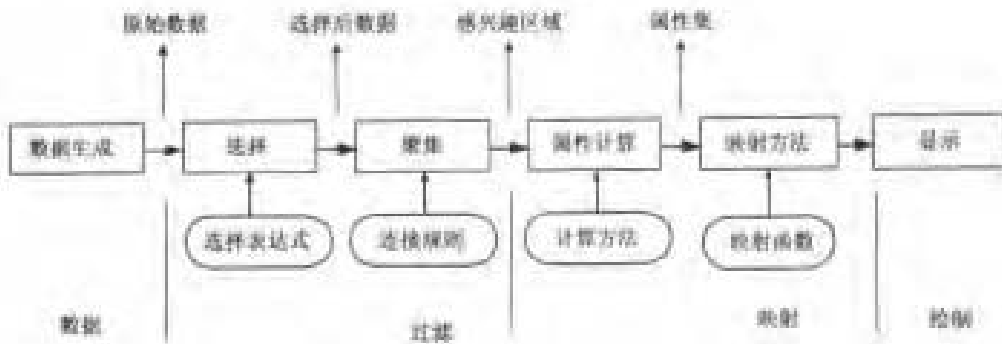


图 41.5.8 可视化的流程

(1) 数据生成 矢量场数据通常由计算流体力学仿真数值模拟产生或者科学实验产生，目前的数据格式有很多种，比较出名的是 plot3d、vvs、nff、cdf 格式。

(2) 数据精炼及过滤 矢量场包含大量的数据，需要经过选择、聚集，或者属性计算后，可以即减少数据量，又可以最大限度地减少有用信息的丢失。需要对数据进行消除噪声、参数域变换和法向计算等。

(3) 数据映射 将处理后的数据转换为可以绘制的几何因素和属性，或者转换成其他感官可以认知的信息，如声音等。映射是整个可视化系统的核心，它主要包括如何用形状、光亮度、颜色以及声音等属性来表示原始数据中人们感兴趣的性质和特征。

(4) 绘制 将产生的几何因素或者声音音符转换为可供显示的图形、声音，图形所用的方法是计算机图形学的基本技术，如透视变换、光照计算、隐藏面消除等。声音显示提供被人感知的、乐于接受和可理解的声音流，经常需要将声音转变为乐音。

### 3.3 矢量场特征可视化

目前工程计算的数据规模呈指数级增长，矢量场更是蕴含惊人的数据量和维度，显示大规模矢量数据蕴涵的所有信息非常困难。虚拟现实系统虽然具有强大的计算能力和图形显示能力，但是即使最高档次的计算机也不能应付目前的工程需要。虚拟现实要求很高的沉浸感、实时的交互性，延迟会破坏虚拟现实营造的真实感，因此基于特征的可视化显得尤为重要。基于特征的可视化是指对计算数据作子集选择、结构分析、特征提取，显示典型特征、关键结构、局部变化或用户感兴趣区域。目前最常用的特征可视化技术是基于拓扑结构分析的，斯坦福大学的 J. Helman 和 L. Hesselink 首先提出基于拓扑结构来揭示、分析矢量场结构的理论和技术，指出矢量场拓扑由关键点、连接关键点的积分曲线和曲面组成。下文将基于拓扑结构的方法，分析二维和三维拓扑结构的关键点分类方法，对矢量场进行特征提取，并且提出了以关键点为种子

点的流场生成策略, 针对涡特征的几何和物理特征综合提取方法。

### 3.3.1 基于拓扑结构的矢量场特征可视化

#### (1) 二维矢量场拓扑结构方法

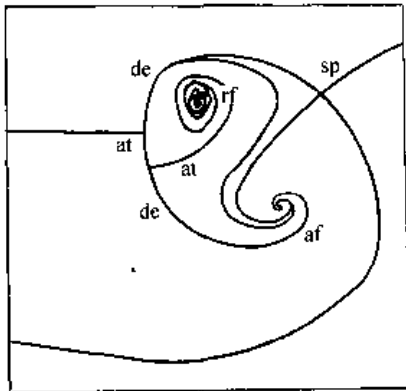


图 41.5-9 二维拓扑结构  
at—聚点 de—排斥点 rf 排斥聚点  
af—吸引聚点 sp—马鞍点

矢量场中的关键点是点特征, 关键点本身可以看作一条无限短的流线或者没有厚度 (两个种子点无限接近) 无限窄的流面, 流线、流面在关键点开始或结束。如图 41.5-9 所示, 拓扑结构包括关键点和在这些关键点特征方向上出发的流线。

基于关键点理论, 关键点是矢量的所有分量都为零的点, 首先将矢量场中所有速度为零的点找出来, 然后根据关键点附近速度场的特性对关键点进行分类。

关键点附近速度场的特性可由其速度矢量对位置矢量的偏导数矩阵  $\nabla V = \frac{\partial V}{\partial X}$  决定, 其雅克比矩阵 (Jacobian) 是一个  $2 \times 2$  的矩阵:

$$J = \left[ \frac{\partial(u, v)}{\partial(x, y)} \right]_{x_0, y_0} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_{x_0, y_0} \quad (41.5-7)$$

关键点的分类根据式 (41.5-7) 两个复数特征值  $(R_1 + iI_1, R_2 + iI_2)$  决定, 见表 41.5-4。一个矢量场的拓扑由关键点和连接关键点的积分曲线和曲面组成。

表 41.5-4 二维拓扑的关键点分类表

关键点名称	特征值	图示
马鞍点	$I_1, I_2=0$ and $R_1 \times R_2 < 0$	

(续)

关键点名称	特征值	图示
排斥点	$I_1, I_2=0$ and $R_1, R_2 > 0$	
聚点	$I_1, I_2=0$ and $R_1, R_2 < 0$	
排斥聚点	$I_1 = -I_2 \neq 0$ and $R_1 = R_2 > 0$	
吸引聚点	$I_1 = -I_2 \neq 0$ and $R_1 = R_2 < 0$	
同心圆 (流的形态由流场邻域的高阶分量决定)	$R_1 = R_2 = 0$	

#### (2) 三维拓扑结构方法

考虑如下的线性系统:

$$\begin{cases} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{cases} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{cases} x_1 \\ x_2 \\ x_3 \end{cases} \quad (41.5-8)$$

那么式 (41.5-4) 特征方程  $\det(A - \lambda I) = 0$  可以写为:

$$\lambda^3 + P\lambda^2 + Q\lambda + R = 0 \quad (41.5-9)$$

有:

$$P = -(a_{11} + a_{22} + a_{33}) = -\text{tr}A \quad (41.5-10)$$

$$Q = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \quad (41.5-11)$$

$$R = - \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -\det A \quad (41.5-12)$$

方程 (41.5-9) 是一个三次方程, 其根可以表示为:

$$D = 27R^3 + (4P^3 - 18PQ)R + (4Q^3 - P^2Q^2) \quad (41.5-13)$$

式中根据  $D$  的取值范围有:

$D > 0$ , 方程有 3 个不等的实根。

$D = 0$ , 方程有三个实根, 其中两个相等。

$D < 0$ , 方程有一个实根, 一对共轭复数根。

三维拓扑结构的分类相对比较复杂, 不同分类子空间内所包括的关键点, 决定了其特性。实际在可视化过程中, 需要将其关键点分类成特定的模式, 如表 41.5.5, 以减小分类计算代价。

表 41.5-5 三维拓扑的关键点类型

关键点名称	条件	图示
对称源 星节点	$3Q = P^2$ 并且 $27R = P^3$ $k = -P/3$	
星型节点	$D > 0, Q > 0$ 同时 $PR > 0$ 有三个相同符号 的实数特征值	同上
马鞍点	$D > 0, Q < 0$ 同时 $PR < 0$ 有三个不同符号 的实数特征值	
鞍点	$D < 0$ , 一对共轭 复数	
中心点	$D < 0$ 同时 $PQ - R = 0$ 两个复数特征值 的实部为 0	
排斥点	$R = 0$ 至少有一个根为 0, 这种情况类似二 维的拓扑结构	

### 3.3.2 种子点选取和布置方法

#### (1) 种子点选取

矢量场可视化主要有两个问题: 矢量场的特征和矢量场表达的清晰程度, 而这两个主要问题都和种子点相关。种子点是积分流线的初始点, 对流线的形态和空间分布影响很大, 如果种子点选取不合适会导致流场的重要特征丢失。种子点的布置不合理, 也会导致流线相互缠绕、聚集一团 (Clutter), 使得流场混乱不堪。

通常在规则的网格节点上选取种子点, 然而种子点作为粒子的靶点不一定就在网络上, 而是在空间某一个位置, 在网络上选取种子点, 可能使得流场中的流

线没有经过马鞍点、排斥重点等关键点, 弱等特征被丢失。为了跟踪流场中的所有特征, 只能增加种子点个数, 流线缠绕非常难以预测, 可通过重采样减少流线密度, 但是仍然不可避免流线的混乱问题。

Image-guided 方法主要利用在流场中生成均匀的流线, 以解决流线的混乱问题, 但是却忽略了流体的特征, 用这种方法在二维流场中, 可能效果不错, 但是对于三维空间流场, 整个流场中流线的均匀布置, 很难去评价。故种子点选择原则是不从网格节点上选择种子点, 而是通过对整个流场搜索关键点, 以关键点作为种子点, 保证流场中每一个关键特征都被跟踪到。

#### (2) 种子点放置方法

正如上文所阐述, 种子点的布置影响到流线的状态, 种子点的选取和布置应该满足三个条件:

1) 覆盖率 可视化后的流线形成的图形应该描述矢量场所有有意义的特征, 不应该丢失一些特征, 同时流线应该充满整个流场, 简单的加大流线的密度是不可取的, 大量的流线缠绕在一起, 反而隐藏和混淆了矢量场的特征。

2) 均匀性 流线应该尽量分布在整个流场中, 这确实是一个挑战性的问题, 因为流场中的种子点位置可以控制, 但是通过这些种子点生成的流线, 其行为却不可预先知道。

3) 连续性 尽量采用较长的流线, 以显得整个流场的连续性, 然而流线越长会使得流场的混乱感加强, 这里必须考虑到均匀性和连续性的平衡问题。

### 3.3.3 矢量场特征可视化实例

以 NASA Blunt Fin 数据进行矢量场特征可视化比较。气流通过一个钝边缘的凸起平板时, 自由气流平行于平板, 在凸起边缘的起始部分有一个非常弱的涡, 图 41.5-10a 是采用普通种子点和 image-guided 方法可视化结果, 看不到涡; 图 41.5-10b 为 NASA FAST 系统的特征可视化, 看出在底部有明显的涡, 利用种子点布置的方法和 FAST 可视化后结果类似, 且是效果更好, 如图 41.5-10c。

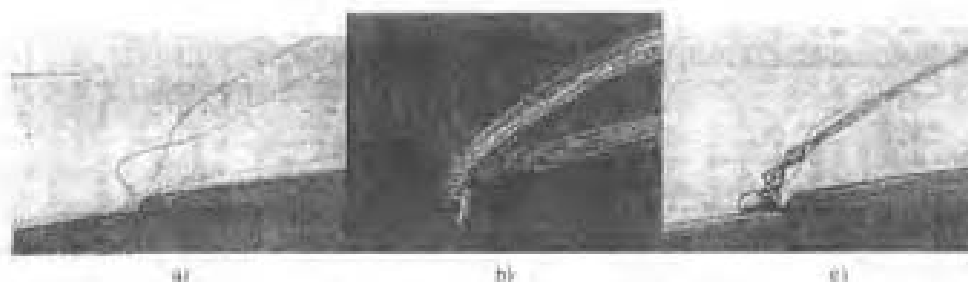


图 41.5-10 基于种子点组织的可视化对比

对比结果可以看到,通用的种子点方法和采用 Image-guided 方法,流场的均匀性得到很大改观,连续性也很好,但是流场的特征体现的不明显,比如马鞍点、吸引聚点不突出,丢失了关键特征。种子点布置的结果满足了覆盖性、均匀性和连续性等特性,矢量场的特征非常明显。

#### 4 矢量场数据的沉浸可视化关键技术

沉浸虚拟现实 (Immersive VR) 给人一种身临其境的感受,人融人在计算机创建的虚拟世界中。当人们沉浸在一个大的视野中,伴随三维声音、带力反馈的交互、逼真的表面纹理和材质,有助于人们对环境的感知和对计算机提供的上下文信息进行理解,辅助空间判断等。

高速计算环境的成熟使得研究人员有更大的空间对复杂工程进行性能仿真,许多复杂的现象因此能够获得研究,尤其是对复杂流场的计算成为可能。科学计算可视化则在三维矢量场的探索使科学家和工程师在洞察工程问题时看到了曙光。科学计算可视化本身不是终点,而是整个科学任务的一部分,它将复杂的数据进行可视化,用来寻找模式、特征、关系、异常和相似形等问题,使原来不易观察的现象有了直观的观察方式。然而工程产生数据的能力超过理解数据的能力,这种差距正在扩大。因此需要寻找一种办法让人参与在可视化的循环中 (Human-in-the-loop Visualization)。计算机在仿真、数据过滤、数据精简上占有优势,而人具有高度发达的模式识别能力,可以看透纷繁复杂数据中的特征与异常。

融合虚拟现实技术和科学计算可视化,称为沉浸可视化 (Immersive Scientific Visualization),当观察者进入计算可视化后的虚拟环境,可形成良好的沟通环境;自然、直觉的可视化方式帮助人更清楚了解事物的内部特性。通过使用多种辅助视图来提供全局数据的上下文提示,可以更自然更快捷地理解三维或多维数据,以人为中心来判断三维空间将变得自然,可以识别和理解三维结构,同时立体的三维深度信息使得矢量的空间判断非常容易。

尽管科学计算可视化的研究很早,但是可视化技术与计算机图形技术的发展并不同步。早期认为点、线等几何图标因为其指向的二义性以及图标的相互重叠聚集,无法清楚表达矢量场,然而随着虚拟现实技术、人机交互技术的逐渐成熟,点、线图标的表达被实践证明非常适合在虚拟环境下表达矢量场。立体显示技术将指向的二义性和图标重叠比较好的解决,相反以纹理映射的矢量场表现方法却受到限制,究其原因,是因为视觉的沉浸依赖图形深度暗示,线积分卷积等方法,

虽然较好表示了某个截面的矢量场结构,但是只能看到其外部,不能透视其内部。正是在这种背景下得到启发,研究融合虚拟现实的矢量场可视化的最佳表达方法,如流线、流管、流面、流球以及三维实体图标。

#### 4.1 适合各种网格单元的自适应步长快速流线生成算法

##### (1) 流线生成基本方法

1) 定义:在流场  $u(x, t)$  中,一条线上所有点的瞬时速度都与该线相切时,该线称为流线。

2) 流线可以通过求解下列常微分方程得到

$$\frac{dx(t)}{dt} = u(x(t)) \quad (41.5-14)$$

式中  $x(t)$ ——空间位置;

$t$ ——时间变量 (pseudo-time) 为积分变量,不是真正的时间。其解为:

$$x(t) = x(0) + \int_0^t u(x(t)) dt \quad (41.5-15)$$

3) 生成流线的步骤 (图 41.5-11) 为:

在物理空间选择初始点。

寻找初始点相关的单元。

将该点转换到计算空间中。

沿粒子路径跟踪下一个点,直到流线结束。

##### (2) 物理空间和计算空间的转换

物理空间是矢量场数据定义的空间,矢量场数据通过计算流体动力学 (CFD) 计算生成。CFD 是利用有限单元方法,矢量场数据在物理空间大多分布在不规则的网格结点上,计算空间是为了计算方便而人为定义的空间,其网格是正交的规则网格,通过建立物理空间和计算空间的转换关系,使得可视化所需的计算在计算空间中完成,大大提高可视化速度。

根据有限元理论,设定物理空间一点  $(x_1, x_2, x_3)$ ,在计算空间中为  $(\xi, \eta, \zeta)$ ,可以得到:

$$\begin{aligned} \Delta x_1 &= \frac{\partial x_1}{\partial \xi} \Delta \xi + \frac{\partial x_1}{\partial \eta} \Delta \eta + \frac{\partial x_1}{\partial \zeta} \Delta \zeta + R_1 \\ \Delta x_2 &= \frac{\partial x_2}{\partial \xi} \Delta \xi + \frac{\partial x_2}{\partial \eta} \Delta \eta + \frac{\partial x_2}{\partial \zeta} \Delta \zeta + R_2 \\ \Delta x_3 &= \frac{\partial x_3}{\partial \xi} \Delta \xi + \frac{\partial x_3}{\partial \eta} \Delta \eta + \frac{\partial x_3}{\partial \zeta} \Delta \zeta + R_3 \end{aligned} \quad (41.5-16)$$

式中,  $R_1, R_2, R_3$  为二阶无穷小,设矩阵  $J$ :

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial \xi} & \frac{\partial x_1}{\partial \eta} & \frac{\partial x_1}{\partial \zeta} \\ \frac{\partial x_2}{\partial \xi} & \frac{\partial x_2}{\partial \eta} & \frac{\partial x_2}{\partial \zeta} \\ \frac{\partial x_3}{\partial \xi} & \frac{\partial x_3}{\partial \eta} & \frac{\partial x_3}{\partial \zeta} \end{bmatrix} \quad (41.5-17)$$



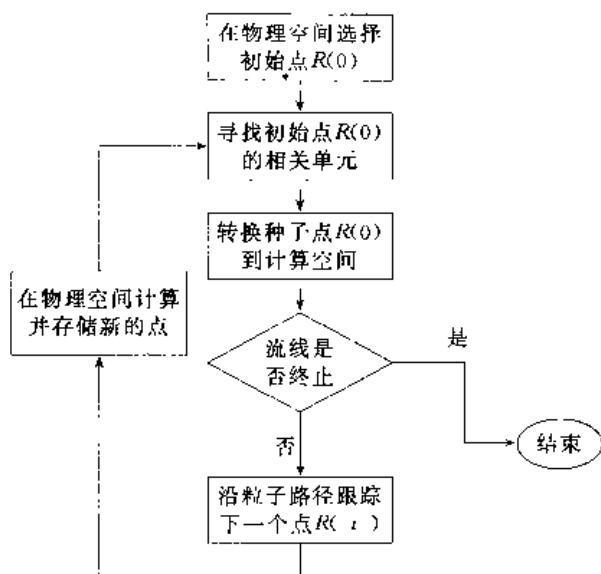


图 41.5-11 流线生成流程图

则式 (41.5-16), 可以写为:

$$\mathbf{x} = \mathbf{J} \cdot \boldsymbol{\xi} + \mathbf{R} \quad (41.5-18)$$

其中:  $\mathbf{x} = [\Delta x_1, \Delta x_2, \Delta x_3]^T$

$$\boldsymbol{\xi} = [\xi, \eta, \zeta]^T$$

$$\mathbf{R} = [R_1, R_2, R_3]^T$$

### (3) 网格单元的插值形函数

然而不同有限元软件, 其使用的网格单元不尽相同, 目前最常见的网格单元是四面体、五面体和六面体等。现有的流线积分方法, 都是在六面体单元中计算, 没有考虑到所有的单元。根据有限单元方法, 可推导出四面体、四棱锥、三棱柱等单元的插值形函数, 从而做到流线积分兼容多种单元格式。

1) 四面体单元 如图 41.5-12 所示, 四面体单元从物理空间到计算空间的转换关系, 有:

$$\begin{aligned} f_1 &= 1 - \xi - \eta - \zeta \\ f_2 &= \xi \\ f_3 &= \eta \\ f_4 &= \zeta \end{aligned} \quad (41.5-19)$$

坐标转换函数可以写成:

$$\mathbf{x} = (1 - \xi - \eta - \zeta) \cdot \mathbf{x}_1 + \xi \cdot \mathbf{x}_2 + \eta \cdot \mathbf{x}_3 + \zeta \cdot \mathbf{x}_4 \quad (41.5-20)$$

式中  $\mathbf{x}_i, i=1, 2, 3, 4$ , 为单元的顶点。

2) 四棱锥单元 如图 41.5-13a 所示, 四棱锥单元从物理空间到计算空间的转换关系, 有:

$$\begin{aligned} f_1 &= \frac{1}{4} (1 - \xi - \zeta) (1 - \eta - \zeta) \\ f_2 &= \frac{1}{4} (1 + \xi - \zeta) (1 - \eta - \zeta) \\ f_3 &= \frac{1}{4} (1 + \xi - \zeta) (1 + \eta - \zeta) \\ f_4 &= \frac{1}{4} (1 - \xi - \zeta) (1 + \eta - \zeta) \\ f_5 &= 1 - (1 - \zeta)^2 \end{aligned} \quad (41.5-21)$$

3) 三棱柱单元 如图 41.5-13b 所示, 三棱锥单元从物理空间到计算空间的转换关系, 有:

$$\begin{aligned} f_1 &= (1 - \xi - \eta) (1 - \zeta) \\ f_2 &= (1 - \xi - \eta) \zeta \\ f_3 &= \xi \zeta \\ f_4 &= \xi (1 - \zeta) \\ f_5 &= \eta \zeta \\ f_6 &= \eta (1 - \zeta) \end{aligned} \quad (41.5-22)$$

4) 六面体单元 如图 41.5-13c 所示, 六面体单元从物理空间到计算空间的转换关系, 有:

$$\begin{aligned} f_1 &= \frac{1}{8} (1 - \xi) (1 - \eta) (1 - \zeta) \\ f_2 &= \frac{1}{8} (1 + \xi) (1 - \eta) (1 - \zeta) \\ f_3 &= \frac{1}{8} (1 + \xi) (1 + \eta) (1 - \zeta) \\ f_4 &= \frac{1}{8} (1 - \xi) (1 + \eta) (1 - \zeta) \\ f_5 &= \frac{1}{8} (1 - \xi) (1 - \eta) (1 + \zeta) \\ f_6 &= \frac{1}{8} (1 + \xi) (1 - \eta) (1 + \zeta) \\ f_7 &= \frac{1}{8} (1 + \xi) (1 + \eta) (1 + \zeta) \\ f_8 &= \frac{1}{8} (1 - \xi) (1 + \eta) (1 + \zeta) \end{aligned} \quad (41.5-23)$$

### (4) 流线积分方法

流线方法源于试验性流场的可视化技术, 流线生成主要有两种方法: 基于数值积分的生成方法和基于流函数的生成方法。数值方法生成流线比较方便, 但是存在误差, 包括跟踪算法无法控制的误差, 如收敛性, 截断误差和跟踪算法本身的误差, 如矢量场插值, 数组积分阶数, 步长选择等; 另外数组积分对初始点非常敏感, 积分初始点的选择也会造成流线生成误差。流函数方法生成流线速度最快, 精度最高, 因为不需要积分, 与选取种子点无关, 只需要构造两个流面函数, 然后求交, 即可生成流线。

但是用流函数生成流线, 对流线上点进行粒子跟踪, 生成流管、流面都非常复杂, 所以在本节中用数值积分流线生成方法。只要选定初始位置, 采用数值积分

的方法,一步步跟踪下去,就可以得到流线上点的位置随时间  $t$  的变化曲线。跟踪既可以在计算空间中进行,也可以在物理空间中进行。

目前生成流线的算法,大多采用一阶欧拉,2阶和4阶 Runge-Kutta 方法,欧拉方法虽然速度快,但是不精确。对于一个均匀的矢量场,如图 41.5-14 可知欧拉方法总是发散而 Runge-Kutta 方法非常精确。

为了提高积分的准确性,用4阶龙格库塔(Runge-Kutta)方法,如式(41.5-24),流线的可视化结果如图 41.5-15 所示。

$$x(t + dt) = x(t) + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \quad (41.5-24)$$

式中  $K_i = dt \cdot V(x(t))$

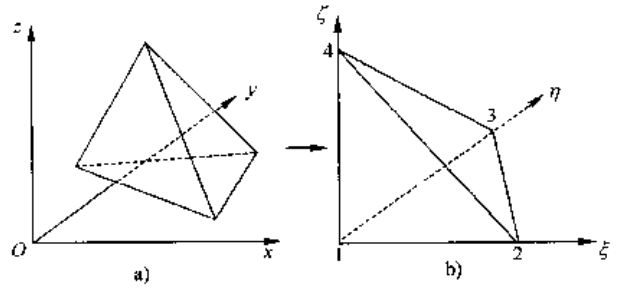


图 41.5-12 四面体物理空间到计算空间转换图

a) 物理空间 b) 计算空间

$$\begin{aligned} K_2 &= dt \cdot V\left(x(t) + \frac{K_1}{2}\right) \\ K_3 &= dt \cdot V\left(x(t) + \frac{K_2}{2}\right) \\ K_4 &= dt \cdot V(x(t) + K_3) \end{aligned} \quad (41.5-25)$$

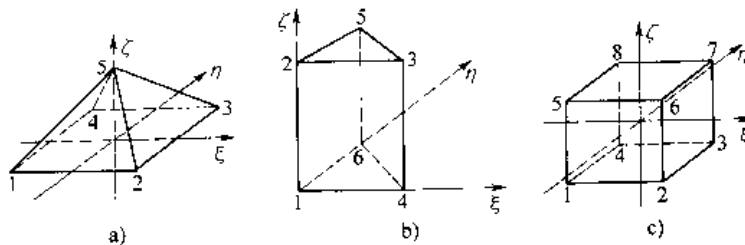


图 41.5-13 计算空间单元类型

a) 四棱锥 b) 三棱柱 c) 六面体

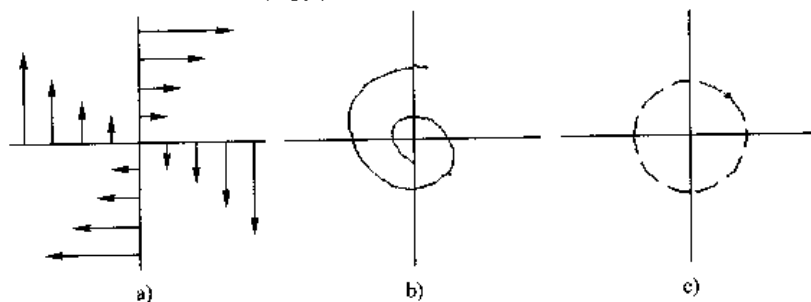


图 41.5-14 均匀旋转流场中欧拉方法和 Runge-Kutta 方法对比图

a) 均匀旋转矢量场 b) 欧拉方法 c) 龙格库塔 (Runge-Kutta) 法

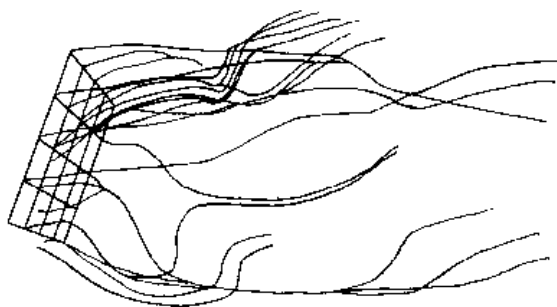


图 41.5-15 流线

(5) 自适应步长流线积分方法

在积分过程中,步长大多是固定步长,但是流线并不是平滑的或者是直线,对于转角较大的地方,如果步

长较大,会使得积分曲线变形,而在相对直的地方,步长较小,会使得计算复杂度急剧增加。Buning 给出了一种计算步长的方法,该方法的步长与单元大小和速度矢量的倒数相关,这种方法需要计算单元的大小,同时联系速度矢量,计算较复杂。借鉴该方法,可以给出自适应步长的改进积分方法,根据微分几何思想,计算流线的转角(winding-angle),步长根据转角按比例缩放步长(如式 41.5-26),尽管本方法的计算时间复杂度较高,但是相比较固定步长的4阶 Runge-Kutta 方法还是快,并且流线准确,反映了流线的细节。

$$s_c = s_{min} + (a - a_{min}) \cdot \frac{s_{max} - s_{min}}{a_{max} - a_{min}} \quad (41.5-26)$$

式中  $s_c$  — 下一个步长;

$s_{max}, s_{min}$ ——预定义步长, 分别表示最大步长和最小步长;

$\alpha$ ——该点转角 (winding angle);

$\alpha_{max}, \alpha_{min}$ ——预定义转角 (winding angle), 表示最大转角和最小转角。

实现如算法 5.1

```

算法 5.1
AdaptiveTimeStep ()
{
  Define stepmin, stepmax;
  Define  $\alpha_{min}$  as windinganglemin,  $\alpha_{max}$  as windinganglemax;
  Choose a initialize value;
  RungeKutta4 (timestep);
  computeWindingAngle ();
  Compute timestep;
  ComputeNextStep;
}
  
```

## 4.2 流面生成算法

最直接的生成矢量面的方法是 Hultquist 提出的通过连接两相邻矢量线上的采样点, 生成三角面片的方法。首先可将流面视为一个参数曲面  $r(x, t)$ ,  $x$  是初始面线  $r(x, 0)$  的参数,  $t$  是构造流面的流线参数, 曲线  $r(x, t)$  是一系列流线。为了构造流线, 先将初始面线  $r(x, 0)$  的参数离散化, 得到一系列的初始离散点  $r_{i0}$ , 然后将从这些离散点上流出的流线离散化, 得到一系列的点  $r_{ij}$ , 将这些点, 线形成的网格连成多边形, 就得到流面  $r(x, t)$ 。由两条矢量线构造矢量面, 则生成流带。这种方法实现简单, 但是生成的图像效果不好, 当矢量场发散、收敛、截断以及绕流时会导致畸形网格, 单纯增加采样密度  $(x, t)$  也效果不好, Helman 等提出如果相邻的两条矢量线发散, 在其间插入一条全长的矢量线。Hultquist 于 1992 年提出 Advancing Front 方法, 在生成流面积分的过程中, 通过相邻矢量线的间距大小, 自动调节采样密度, 当发散时提高采样密度, 收敛时降低采样密度, 其唯一的缺点就是每前进

一步都要判断, 对初始曲线的形状也有所限制, 难以解决矢量面的生成。

而以一条直线为一条边, 取流线上生长出定长的线段, 线段的方向为其速度的法矢, 连接所有的线段端点就构成了另一个边, 来构造流面, 不失为一种更加容易的方法, 如图 41.5-16 所示。



图 41.5-16 流带生成算法

定长的向量旋转角度由下述公式决定:

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{1}{2}(\omega + \dot{S}) \\ \dot{\omega} &= \text{curl}(u^*) \\ \dot{S} &= \frac{u^* \cdot \dot{u}^*}{\|u^*\|^2} \end{aligned} \quad (41-5-27)$$

式中  $\theta$ ——旋转角。

实现如算法 5.2:

```

算法 5.2
CreateRibbonRibbon ()
{
  Do while streamline not end
  Seedpoint, start;
  IntegerLine (seedpoint, 0);
  ComputingNormalVector (seedpoint, x0);
  Repeat next seedpoint;
  ConnectOtheredge (otherpoint [j]);
  Wend;
}
  
```

该方法生成流面不会出现发散或者收敛的情况, 也不需要生成流面的时候调整采样, 因为流面的生成和流线其他点无关, 生成速度快, 流带图形质量高, 但是这种方法生成的流带不能反映流场流线邻域的流动属性和流线邻域的高阶特征, 在大多数的情况下, 这种方法是有效和高速的, 如图 41.5-17 所示生成的流带。

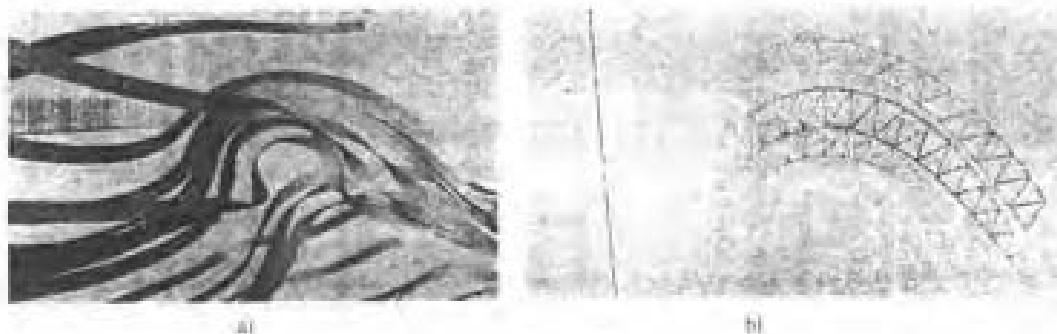


图 41.5-17 流带

### 4.3 流管生成算法

为了充分利用图形学的光照、色调、消隐、透视投影等所提供的三维深度信息，如方向、扭曲等，将三维矢量数据映射成流管，可以有效地避免流线集中时视觉上的混乱现象。流管因为是有半径的实体，所以通过其体特性可以表达矢量场的一些局部属性，如应变、位移以及旋转等。

基于扫描生成流管的方法通常有两种：多边形和椭圆扫描方法，见图 41.5-18 和图 41.5-19。

多边形为正多边形，经常采用的是 5~8 边形，取决于显示的精度，多边形的大小由外接圆的半径  $r$  决定（图 41.5-18a），而多边形的朝向始终垂直于速度方向（图 41.5-18b），显然对于多边形生成的流管，其

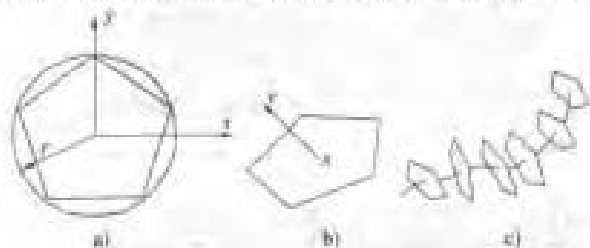


图 41.5-18 多边形扫描生成流管  
a) 多边形 b) 多边形垂直于速度  
c) 多边形沿流管方向扫描

流管的属性由外接圆决定。

而椭圆形成的流管由椭圆的长、短轴决定，见图 41.5-19。长短轴的值可以表示差量的特征向量和特征值：

$$w = \lambda e_i, \quad i = 1, 2, 3 \quad (41.5-28)$$

式中  $e_i$  ——特征值方向的单位向量；  
 $\lambda$  ——特征值。

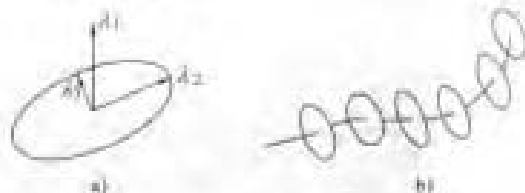


图 41.5-19 椭圆扫描生成流管  
a) 椭圆 b) 椭圆沿流管方向扫描

椭圆扫描面被称为超流线 (Hyper Streamlines)，如果其长轴等于短轴，其扫描面为圆，Shyh-Kuang 等人研究了非结构网格下的流管，他们将流管的半径  $r$  由如下常微分方程定义：

$$\frac{1}{r} \frac{dr}{dt} = \frac{1}{2} \nabla_{\perp} \cdot u^* \quad (41.5-29)$$

式中， $\nabla_{\perp} \cdot u^*$  ——局部紊流 (local cross flow) 的散度，故该流管是变半径的流管。

利用该方法生成的流管如图 41.5-20 所示。

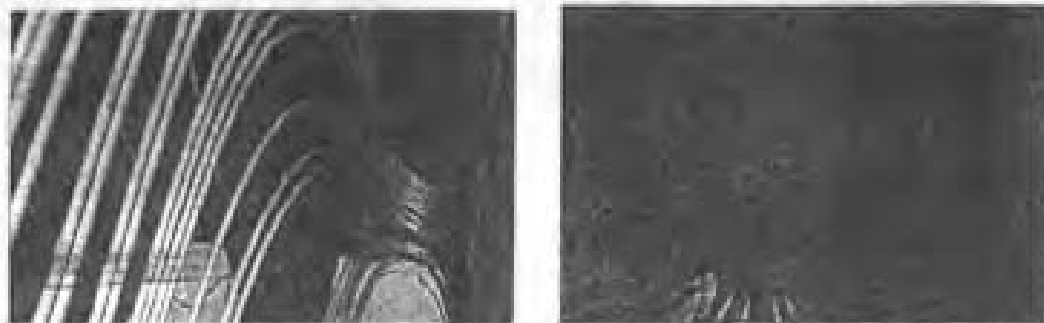


图 41.5-20 流管

### 4.4 三维实体图标

图标法是表达矢量场最简单的方法，虚拟现实技术的出现使得这些看来被舍弃的方法，重又恢复了生命力。De Leeuw 和 Van Wijk 扩展了实体图标，来表述局部矢量的速度、曲率、旋度、散度以及加速度等，如图 41.5-21 所示。实际上，用如此复杂的图标来表达矢量场，不见得会取得很好的效果，因为这些图标用来描述矢量场的细节，在表达整个流场时，会看起来凌乱。用几何属性和纹理来描述矢量的高阶特征，如弱

量、散度等，不是一个很好的方法（在下节中将引入声音来描述高阶特征）。另外，复杂的几何图形生成会引起计算的急剧膨胀，交互性受到影响。通过实验认为，普通的三维实体坐标，如图 41.5-22 所示，在生成负视觉的立体图形时，效果最好，同时深度暗示已经给用户强烈的方向感。

为了验证算法的有效性，用来自 NASA 数据（称为 Lox Post，液体氧流过一个垂直圆柱的仿真数据），可视化结果如图 41.5-23 箭头图标和图 41.5-24 流球所示。

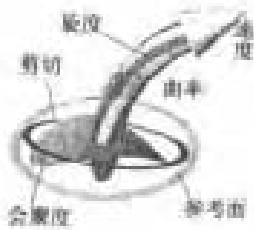


图 41-5-21 流场视觉图标

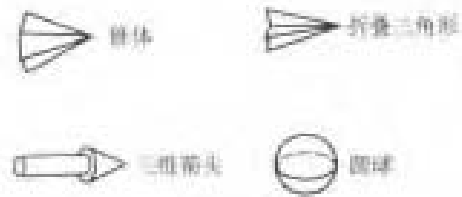


图 41-5-22 简单几何实体图标

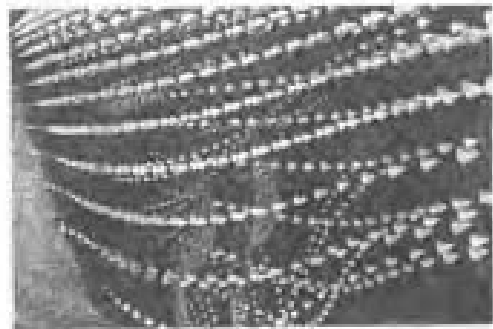
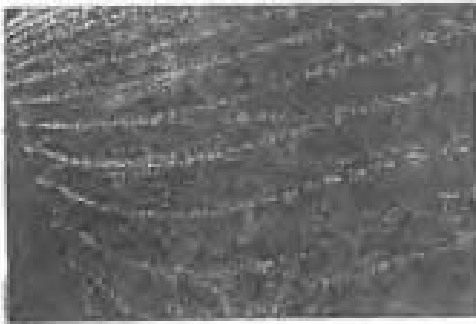


图 41-5-23 箭头图标

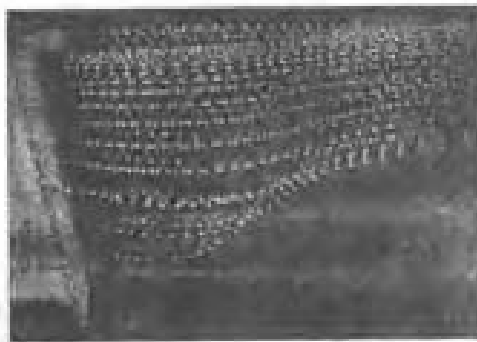


图 41-5-24 流场

### 4.5 颜色映射方法

颜色映射方法其实是普通的标量场可视化技术，就是将标量数据用颜色表示出来。在矢量场的可视化中，颜色映射方法一样是重要的技术，将矢量的大小映射成颜色，采用颜色索引法来完成颜色映射，增加矢量场表达数据的维数。具体颜色对照表如图 41-5-25 所示，映射关系下式。

$$\begin{aligned}
 & a_i < \min, \quad i = 0 \\
 & a_i > \max, \quad i = n - 1 \\
 & i = n \left( \frac{a_i - \min}{\max - \min} \right)
 \end{aligned} \quad (41.5-30)$$

式中  $a_i$ ——矢量的大小，其范围在  $(\min, \max)$  间。

合适的颜色映射需要反复地选择，不但要考虑到真实的颜色还要考虑到颜色的传递含义，如对压力、温度和速度大小采用红—蓝映射模型，大的数值为红色，反之则为蓝。

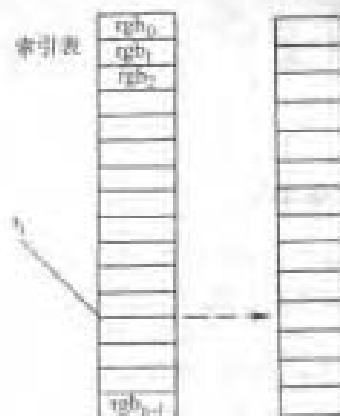


图 41-5-25 颜色对照表

## 第6章 虚拟装配技术

### 1 虚拟装配的概述及其国内外研究简介

#### 1.1 概述

众所周知,产品装配是将零散的部件组装成为完整的产品,它是产品制造过程的最后一个环节。传统的产品装配过程一般是借助于实物模型来完成,如果出现任何一个小的修改都可能导致实物模型的重建,因此这是一个费时、费力的过程并造成了财力、物力的浪费。而面对日益激烈的全球化竞争,要求企业不断缩短产品的生命周期使产品快速响应市场。

虚拟装配(Virtual Assembly)技术正是在这种需求背景下产生的。虚拟装配一般被定义为:无需产品或支撑过程的物理实现,只需通过分析、实验模型、可视化和数据表达等手段,利用计算机工具来安排或辅助与装配有关的工程决策。虚拟装配是一种将CAD技术、可视化技术、仿真技术、决策理论及装配和制造过程研究、虚拟现实技术等多种技术加以综合运用的技术。

传统的CAD系统的装配过程是:设计者逐个调入零部件,利用定位技术,选取相匹配的几何特征并输入几何约束,系统反馈给用户一些约束信息,如一致性检查、碰撞等,系统经过几何约束求解后,零部件就装配在所约束的位置上,同时系统显示装配后的模型,可以演示其装配过程。

然而传统的装配过程,路径用户必须事先确定好,装配者在真实环境中的装配往往非常复杂,这种“爆炸图”式装配并不能给装配设计带来什么,或者说传统的装配仿真根本就不是面向装配设计的,这样零件的可装配性,被隐藏了起来;另外用户采用的是二维的鼠标和输入设备,对零件的装配不自然,指导用户的装配功能非常有限,给用户带来的视觉信息也少。

在虚拟现实环境中的虚拟装配系统,它有三个明显的特点:

- 1) 设计中通过三维输入设备直接对零部件进行三维操作,非常直观而且具有较高的交互性。
- 2) 三维显示设备让设计者可以象在真实世界中一样观察物体。虚拟装配比传统的CAD系统的动态装配,具有更真实感和实用性,更能适应虚拟造型、装配规划,仿真的要求。

- 3) 完全自由的装配过程,通过自然的装配,可以模拟装配者的实际装配过程,这样不仅可以检查可装配性,还可以研究装配过程的人机工程。

国内外对虚拟装配都做了大量的研究,但是目前要使虚拟装配得到普遍使用还有相当长的路要走。

#### 1.2 国内外研究简介

美国Northrop公司的自动飞机机身装配程序(AAAP)是一个图形化的装配模拟环境,它允许设计者通过三维空间中实时操纵零部件对飞机装配中的装配策略、工具应用等进行验证。该系统由立体显示系统、三维鼠标和商用虚拟现实软件等组成,能实现实时碰撞检测、零件穿透预防、装配件的三维配合(Snap)等。Northrop公司将之用于美国空军F18战斗机的改进造型中。

VADE(Virtual Assembly Design Environment)是华盛顿州立大学开发的一个沉浸式虚拟装配系统,在SGI ONYX2工作站上实现。立体视觉由VR4头盔提供,设计者的手的移动可以被Ascension公司的Flock of Bird™所跟踪;手指和手腕的运动由Cyberglove™数据手套监控;Tactools™触觉反馈系统在用户的指尖处产生触觉反馈。该系统中,设计者可以对公差进行评估,选择最理想的装配顺序,生成装配/拆卸路径以及观察最终结果,见图41-6-1。

德国Bielefeld大学的B. Jung等人基于构造工具箱(Construction Kits)的概念建立了一个装配系统CODY。所谓构造工具箱,是指在制造领域中可以重复使用的、具有多重功能的标准零件库。CODY是一个基于知识的、三维交互式虚拟装配系统,它允许设计者在虚拟环境中通过直接三维操作或简单的自然语言命令与系统交互,利用标准的机械零件构造复杂的装配体。在CODY中,物体间可匹配性的知识用于支持诸如装配、拆卸、子装配件的旋转等各种操作。事实上,CODY的一个最主要的贡献即在于它开发了知识表达与推理机制COAR(Concepts for Objects, Assemblies, and Roles),并将之用于建立和维护虚拟装配体模型,从而使用户能够在虚拟环境中利用直观的交互方式设计新的虚拟原型。见图41-6-2。

德国弗朗荷夫工业工程研究所(Fraunhofer Institute for Industrial Engineering)的A. Roessler等断言,目前的虚拟现实系统在精确性上不足以满足虚

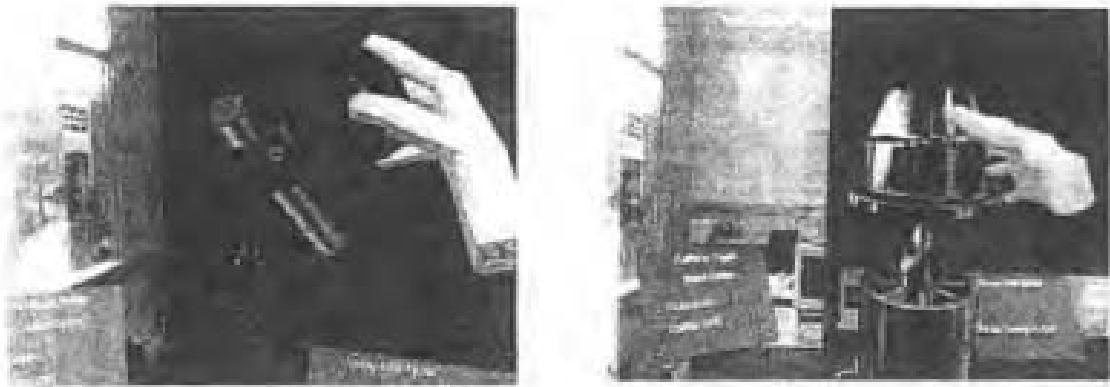


图 41-5-1 VADE 虚拟装配环境

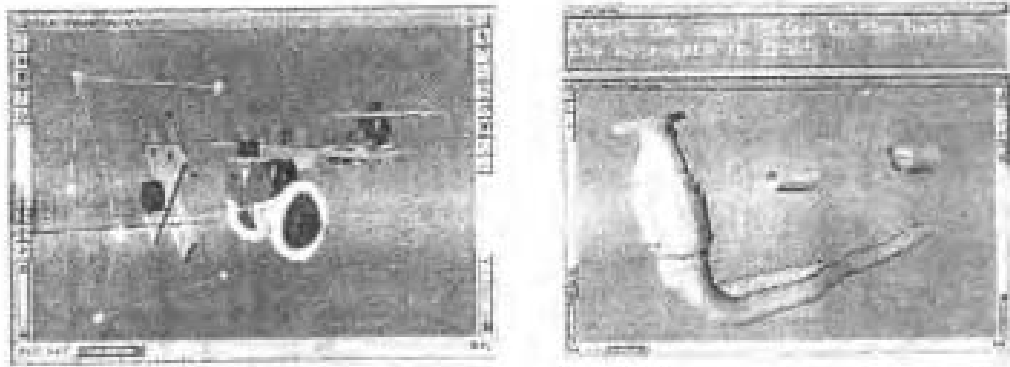


图 41-5-2 CODY 系统

虚拟装配的要求,他们利用空间约束(Space Restriction)算法开发一个帮助设计者在虚拟环境中对零件进行装配和拆卸操作的系统。而B. Antonishek等认为,采用HMD的完全沉浸虚拟环境尽管能够使用户在计算机创造的虚拟环境中体会到真实的现场感,但同时隔绝了用户间的交流,这对于大型设备的装配并不合适,他们基于并行技术公司(Concurrent Technologies Corporation)的Virtual Workbench™和数据手套,在SGI工作站上用WorldToolKit™开发工具开发了一个半沉浸式的虚拟装配系统,见图41-5-3。Virtual WorkBench™是一种水平显示设备,用户戴上立体眼镜后观察的物体就像浮在桌面上一样,更加适合装配,该系统采用模拟人手自然交互方式的双手手势交互,使得诸如对物体抓取、平移、旋转等操作更加自然直观。

目前,国内外不少的研究机构及大型企业在虚拟装配技术研究方面取得了很大的进展。美国华盛顿大学与美国国家标准技术研究局NIST(National Institute of Standard and Technology)合作开发的虚拟装配设计环境VADE(Virtual Assembly

Development Environment);德国Bolefeld大学致力于将虚拟现实交互技术与人工智能技术结合,开发了基于指示的虚拟装配系统CODY。国内的浙江大学CAD&CG国家重点实验室在四面投影虚拟环境CAVE中开发了完全沉浸式虚拟装配原型系统IVAS(Immersive Virtual Assembly System);上海交通大学在PowerWall环境下开发了面向大型复杂零件的虚拟装配系统。

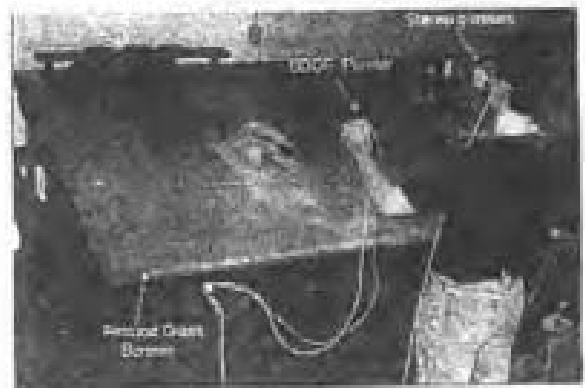


图 41-5-3 Virtual Workbench 平台

## 2 虚拟装配关键技术

### 2.1 虚拟装配模型

#### 2.1.1 传统的装配模型

装配模型是指一组零件经过一系列的装配操作后生成的具有确定关系的模型,是对装配体的一种抽象表达。一般而言,产品的装配模型包括零件的实体信息、零部件之间的装配关系,以及由装配关系所决定的空间位置关系。装配建模的目的就是为了描述装配零件间的关系,因此不论采用什么方法建模,都必须满足两个条件:首先包含零件本身的所有特征约束关系,其次还要包含零件间的所有功能尺寸关系。

零件的实体信息是装配体中各零件、部件实体信息的总和,如点、线、面、精度等几何信息、表面材质、颜色等物理特征。

零件的几何信息建模的表达方式通常有三种:CSG、B-Rep 和 CSG 与 B-Rep 混合表达的方式。CSG 易于表达零件形状抽象信息和零件创建的过程信息,它把复杂的实体看成是由若干较简单的最基本实体经过一些有序的布尔运算而构造出来的;B-Rep 表达模式就是用面、环、边、点来定义形体的位置和形状,它擅长表达零件具体的几何/拓扑信息;而 CSG 和 B-Rep 混合表达模式综合了两种表达方式的优点,即支持多层对象的几何抽象,又便于以交互方式进行基于特征的设计。一般零件建模宜采用 CSG 和 B-Rep 的混合表达模式。

建立在产品装配模型上的装配关系可表示为:  $MR = (MT, ME)$ 。其中 MR 表示零件之间的配合关系。MT 表示零件的配合几何元素,通常指零件间配合的点、线、面几何元素,这些特征类型又隐含了零件的装配或拆卸方向。ME 表示零件之间的配合类型。

装配关系是建立零/部件之间约束的关键。由一般的机械知识可知,产品的装配关系可以分为以下三类:

1) 定位关系 定位关系分为平面贴和、平面对齐、直线对齐、柱面贴和、相切、点面接触等类型。通过零件的定位关系能够描述装配体中零件的空间位置和配合关系。

2) 联接关系 联接关系分为螺纹联接件、键和花键、销、联轴器等类型。联接件一般是标准件,可以预先定义其模型库,设计过程中可以随时提取。

3) 运动关系 零/部件之间的运动关系可分为相对运动和传动两类。相对运动关系包括旋转运动、平面运动等。传动关系包括齿轮传动、带传动、链轮传动和螺旋传动等类型。

目前,装配模型已从图表达的拓扑结构转换为由树表达的层次结构。图 41.6-4 是一个简单的树结构模型。

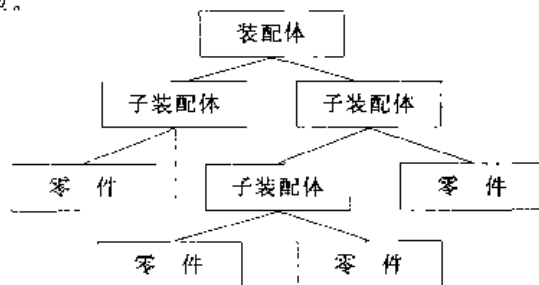


图 41.6-4 装配体的树结构模型

其中树的根结点表示用户最终需要的装配体,非叶节点表示子装配体,子装配体是由零件或子装配体构成,零件在这个树状结构中用最底层的叶节点表示。允许一个零件节点即叶结点有多个父节,表示同一个零件可在同一个装配体中发现多次,系统自动复制零件实例。

虽然树状结构具有操作简便、维护方便、存储量小的特点,但在装配模型上仍存在问题:

1) 信息量不够,难以满足后续装配规划对装配体信息的需求。

2) 装配模型较复杂,建模的难度大。

3) 不能充分利用信息存储结构,造成结构表达的资源浪费。

4) 装配建模技术基于几何的,而工程设计过程中包含了大量的非几何信息,基于几何建模的设计系统不可能支持产品设计的全过程。

而虚拟现实技术为设计者提供了视、听、触觉一体化逼真的设计环境,避免了物理样机的使用。随着虚拟现实技术的日益成熟,虚拟装配将逐步代替主流的传统装配方式,成为重要的应用手段。

#### 2.1.2 虚拟环境下的装配建模

虚拟环境具有很大的逼真性,因此对虚拟装配模型提出了更高的要求:

1) 装配模型的信息高度集成。

2) 装配模型的结构表达便于信息的高度提取。

3) 由于虚拟环境是现实世界的真实映射,对装配体的物理特性信息有特别的要求。

目前,虚拟环境下模型的表示方法有以下几种:

(1) 分层次的零件信息表达

将 CAD 系统中的零件设计信息以中性文件的形式进行存储,然后在虚拟环境中通过读取中性文件来获得零件的设计信息。刘振宇通过研究将零件信息分为零件层、特征层、几何层及显示层,并通过数据映射与约束映射,实现零件信息的层次间关联。分层次的零



件信息模型在满足 VR 交互实时性的同时,保存了 CAD 信息的完整性,并有利于建模过程中根据不同的任务对不同层次的零件进行操作。

其中的零件层以零件为基本节点。零件节点信息包括零件的标识、名称、形状包围盒、物理属性、运动属性、方位属性、显示属性。零件层中零件间的关联体现为各属性参数间的约束关系。特征层以特征为基本节点。特征节点描述的信息包括特征的类型、特征的方位及特征的参数。特征节点间的关系主要体现为特征间的父子关系与约束关系。几何层以面为基本节点。面的节点信息包括该面的 NURBS 曲面描述以及组成该面的环、边、点等边界信息。显示层以三角形小面片为基本节点,记录了组成零件的各小面的顶点坐标、顶点法矢、面片颜色及纹理信息。显示层数据主要用于虚拟环境中模型的显示控制与装配过程中的碰撞检测。

## (2) 使用装配树结构

由于普通的装配树结构仅能表达装配层次信息,而如拓扑信息、约束信息等很多有用的信息却无法表示出来。因此从使信息全面、提高信息的搜索效率方面对模型加以改进;增加装配体部件的标识,增加上层装配体同下层成员的联系,在结构表达中,装配体的组成成员以链表的形式记录便于增加和删除,改进装配体约束信息的存储结构。基于装配体的树状模型,在成员节点上增加装配体的约束信息,同样以链的形式加以存储,从装配体的第一层每一个成员都指向与自身定位相关的装配约束链表,在约束链表中,每一个节点包含一个地址域,该地址指向与头节点成员发生约束关系的同层其他成员节点。最后增加装配体的拓扑关系信息,由于装配工艺路径规划和公差分析都以拓扑信息为操作对象,因此每一个装配体的结构又能引出组成部件的连接拓扑信息。

而要在一个虚拟环境中规划出符合实际要求的装配序列和路径,不仅要考虑装配体零件的几何特性,还要考虑到产品的物理属性。一般而言,装配体的物理属性可分为两类:一类是基本物性,它是指装配体最基本的属性而且一般在工程设计中不再进行分解,如密度、体积、重心位置、硬度、材质等,它们的值在几何建模时就可以确定;另一类是二次物性,一般由基本物性和装配体几何信息共同推导出的属性,如质量、惯性矩、惯性张量、摩擦因数等等,这些信息对于以物性分析为主要特征的虚拟装配设计尤为重要。关键问题是如何在装配树结构上加入物性信息。

由于现实世界中的装配体的复杂程度难以估计,如果每个零件均携带物性结构信息,则会造成加入装配体的树状结构的规模庞大,不便于存取、操作。另一方面,并非每个零件的物性信息在装配规划中都会被

利用,而且在装配体中还存在大量的被重复使用的零件,它们的物性完全相同,因此没有必要把信息种类和信息量均很丰富的物性特征完全加入到整个装配体结构,可采用一种动态存储机制。即在系统中建立一个统一的基本物性库,主要记录装配体的基本物性。

目前,CAD 系统与虚拟装配系统之间常用的信息转换方式包括:

1) 直接利用 CAD 系统提供的标准信息转换接口进行产品模型的转换,其中,VRML 是最常用的格式。这种信息转换方式虽然直接方便,但往往会造成零件拓扑信息、精确的几何信息即几何约束、设计意图、设计历史等信息的丢失。

2) 在 CAD 系统中进行二次开发,将 CAD 系统中零件的设计信息用可扩充的中性文件形式进行表达。现有的三维软件大都有草图、零件设计、部件装配设计、总装设计模块。如 MDT、UG、Pro/E 等,可用来生成三维装配模型,可用现有的 CAD 系统提供的二次开发工具,如 MDT 提供的 Object ARX 类库供 Visual C++ 开发;Pro/E 具有 Developer 与 toolkit 开发模块设计接口。这种方法的优点是能够根据虚拟装配的需要自定义中性文件的格式与内容,但需要对每一个 CAD 系统定义相应的数据转换接口。

在虚拟装配环境下获取产品装配关系主要有两种途径,一种是由设计者交互指定装配体中零件间的装配关系。首先指定待装配零件与已装配零件中的几何体素或称之为目标实体,然后指定待装配零件与目标实体之间的装配关系。如果待装配零件的数量比较大,那么设计者的交互设计工作量将会十分庞大。另一种是由系统自动识别零件间的关系,即系统根据设计者的交互操作,实时地捕捉设计者的装配意图,达到能够识别并建立零件间的装配关系。目前实现装配关系自动识别的方法有:Fa 提出的基于直接三维操作和约束的实体造型方法,它通过几何约束自动识别与允许运动推理来实现装配操作的精确定位。另外还有刘振宇<sup>[21]</sup>等提出的基于装配语义的方法,根据装配语义蕴含的装配配合关系、装配层次关系、装配动作、装配顺序、装配规则与装配参数等信息引导几何约束识别,能有效地提高几何约束识别的效率与准确性。

## 2.2 虚拟装配过程

进行虚拟装配操作时,规划者佩戴好数据手套、头盔式显示器等虚拟环境的输入输出设备后,就可以通过特定的操作界面获得存放在数据库中的关于装配零部件、工具的信息,还可以通过操作界面以各种形式交互地操纵装配产品的各个部分。

从虚拟装配的过程模式来看,以装配仿真为核心

的虚拟装配与以“变装为拆”的装配规划为核心的虚拟装配是当前两种主流的虚拟装配过程模式：

#### (1) 以装配仿真为核心的虚拟装配过程模式

首先，在商用CAD系统（如Pro/E）中建立产品的装配模型。然后，在虚拟环境中模拟产品的实际装配过程，分析装配过程中零部件的运动形态，检查装配过程中的干涉，评价装配过程中的人机因素，从而验证与改进产品的装配性能。这种虚拟装配过程模式实际上是对CAD系统的装配仿真技术。

#### (2) 以“变装为拆”的装配规划为核心的虚拟装配过程模式

首先，在商用CAD系统中建立产品的装配模型。然后，对CAD系统生成的装配模型（包括零件模型及其在装配体中的最终位置、装配约束关系、装配层次关系）通过数据接口转化到虚拟环境中。对虚拟环境中的装配模型进行人工拆卸操作，根据拆卸过程中产生的拆卸顺序与拆卸路径并基于“可装即可拆”的假定，获取产品的拆卸序列与拆卸路径。最后，在以“变装为拆”的装配规划为核心的虚拟装配过程模式中，主要利用CAD系统装配建模提供的装配约束关系进行运动导航，即虚拟拆卸过程中根据零件所受的装配约束拆卸零件的可自由移动方向，并将零件的初始拆卸方向强制为其可自由移动的方向，直到零件与其约束零件完全脱离。

根据零部件之间的装配关系和约束条件，在虚拟环境下进行设计组装，并进行相应的检验，从而对设计进行分析评价，对不合理的设计进行修改达到设计优化的目的。虚拟装配过程包括产品结构分析、装配关系确定和产品组装，各类检验进行设计分析评价及设计修改和装配效果检验等。

虚拟装配过程中首先进行装配过程规划，它必须包含足够的过程信息来支持过程选择决策，并获得更详细的过程定义。如构成顺序（有可选择性）、特征数据、装配轨迹、零件方位以及公差数据等。同时虚拟装配要考虑模型的物理学属性，比如，在现实世界进行装配操作，用木头或泥土等建模材料建造物理模型时，装配零件的惯性属性不同于真正零件的属性；而在虚拟环境中，装配部分的惯性属性将不再起作用，所以应将速度、加速度等合并起来，使虚拟装配过程得到加强。另外，可利用听觉来驱动对物体的各种操作，因此模型应该是一物理模型，应考虑物体的重量、大小、外形、摩擦力等物理特征。

其次在虚拟装配过程中，还需要考虑产品零部件之间的层次关系、装配关系及约束关系等。判断装配的可能性和通畅性，确定各对象之间的公差分配，选择好的装配方法和装配路径（装配可达性）。因此装配系统

如何精确表示产品的装配过程是一个必须解决的问题。需要记录各部件在空间中的运动轨迹，这样的轨迹形成软区域，需要实现创建、存储和显示这些软区域的方法，将轨迹送到CAD系统之前，需要整理零件的装配轨迹，轨迹数目要减少，轨迹需要优化。

虚拟装配过程应形成装配工艺，即把产品在VR系统中的装配描述映射到装配工艺中。虚拟装配过程将产生一些有用的信息（装配顺序、装配轨迹）等，这些信息应被送到CAD系统中以便于进一步修改。

## 2.3 碰撞检测

### 2.3.1 虚拟装配中的碰撞检测

在虚拟装配过程中，碰撞检测技术是涉及到的一个相当关键的技术。众所周知，在装配过程中不可避免地会出现碰撞现象，进行碰撞检测。碰撞检测包括静态干涉检测和动态碰撞检测。其中静态干涉检测又称为可装配性检测，它检查零件的包容体的干涉情况。动态碰撞检测又称为可达性检测，它可通过由装配路径等信息形成的虚拟体的方式转化为静态干涉检测而实现。碰撞检测问题关键之一是如何减少计算量。

碰撞检测算法用于两个方面。首先在用户引导的拆装过程之前，用精确的碰撞检测算法检查零件间是否发生干涉，进行静态干涉检查。再就是当用户引导的拆装开始后，碰撞检测算法用于避免零件在拆装的过程中与其他零件的碰撞。

在虚拟现实系统中由于系统苛刻的实时性限制对碰撞检测算法有非常高的要求，目前常用的是精确碰撞检测算法，该算法可以有效地处理任意形状的多面体间的碰撞问题。为装配体在世界坐标系中建立一个包围盒。当正在拆装的零件仍位于装配体包围盒内部时，它需要与其他仍未拆卸的零件进行碰撞检测；一旦零件被完全移出装配体包围盒，即可对该零件的装配检测进行精简处理。同时对装配的空间和装配的零件的状态也进行限制，因为可以认为装配体拆卸的零件尽管放在虚拟环境中，但不必对他们进行碰撞检测。有几种方法可达到如此的要求；比如设定装配状态，对于装配在一起的零件其空间装配状态是真；而拆开的零件，状态为假。或者利用零件距离装配体的远近来判断是否要进行碰撞检测，假定工作台位于 $Y=0$ 的平面上，正在拆卸的零件 $P$ 与工作台间的碰撞检测简化为只需判断零件 $P$ 的多面体表示中，其最低的顶点的 $Y$ 值是否小于零即可，若小于零，则存在碰撞，反之，零件与工作台无碰撞。

### 2.3.2 公开算法软件包

成熟的碰撞检测算法有很多，比如层次包围盒、

AABB 树算法,集合操作算法等,读者可以查阅相关资料。目前有一些成熟碰撞检测算法的软件包在网上公开,可以自由使用,它涵盖了目前主要的碰撞检测算法,可以根据应用选择使用。

#### (1) RAPID

网络地址: <http://www.cs.unc.edu/~geom/OBB/OBBT.html>

RAPID 用于精确检测两个物体间碰撞的库,其采用 OBBTree 算法,输入的模型为“多边形汤”,即物体模型仅是一组无结构的多边形,不包括任何连接信息和拓扑约束。RAPID 比较适合检测两个距离非常近的物体间碰撞。

#### (2) I-Collide

网络地址: <http://www.cs.unc.edu/~geom/ICOLLIDE.html>

I-Collide 用于精确检测环境中包含多个凸多面体的碰撞检测情况,输入模型要求是凸多面体,它利用物体的“凸性”来加速。它采用包围盒排序法找出可能发送碰撞的物体对,利用 Lin-Canny 算法来精确检测两个物体间的碰撞。

#### (3) V-Collide

网络地址: <http://www.cs.unc.edu/~geom/VCOLLIDE.html>

V-Collide 用于大的动态环境中的碰撞检测,输入模型可以是“多边形汤”,它结合了 I Collide 的高层处理多个物体碰撞的包围盒排序法和 RAPID 的处理两个物体的碰撞检测算法,比较适合于包含大量静态或动态多面体的场景,并允许动态地增删物体。

#### (4) V-Clip

网络地址: <http://www.merl.edu/projects/vclip/>

V-Clip 可以跟踪计算多个多面体间距离的算法库。它对 Lin-Canny 算法作了改进,可以处理一对凸多面体,或一对表示成凸多面体层次的非凸多面体。除了可以计算出两物体间的距离外,还可以报告两物体的穿透位置,及发生穿透的物体对间的近似距离。

#### (5) Enhanced GJK

网络地址: <http://www.comlab.ox.ac.uk/oucl/users/stephen.camcron/distance.html>

Enhanced GJK 可以跟踪两凸多面体间距离的算法库,为获得最佳性能,它要求输入模型中提供一个包含两个物体的所有边的表。

#### (6) SOLID

网络地址: <http://www.win.tue.nl/cs/tt/gino/solid/index.html>

SOLID 用于多个物体间碰撞检测的算法库,采用

AABB 树算法,输入物体模型可以是多个多面体或“多边形汤”,适用处理可变形物体的碰撞检测。

#### (7) QuickCD

网络地址: <http://www.ams.sunysb.edu/~jklosow/quickcd/QuickCD.html>

QuickCD 用于两个物体间碰撞检测的算法库,其中一个场景指场景中的静态部分,另一个指在场景中飞行的一个物体,采用 K-dop 算法,输入模型是“多边形汤”,适用处理一个物体在大环境中飞行,比如场景漫游系统。

## 2.4 虚拟装配路径规划和仿真

虚拟装配路径的规划和仿真的目的在于对装配体进行可装配性分析,给出一个可行的装配方案,如装配顺序、各零件的装配路径等信息。基于“可拆即可装”的假定,通过用户引导的拆装过程,来获取零件的拆装信息,然后再将拆装信息予以反向推演,即可获得与零件相关的装配信息。

进行零件拆卸时,系统记录了每一个采样点,要在每一个采样点上零件都与其他零件不发生碰撞,这样记录了零件的全部位置信息。但当利用三维鼠标、数据手套直接操纵零件对其进行拆卸时,具有相当的随意性,显然带来了数据的冗余,在系统进行求解时是不经济的。这时需要对路径进行优化。

#### (1) 合法采样点

指的是记录下来的零件拆卸路径的采样点满足下列条件的点:

零件在装配体的初始位置,即拆卸路径的第一个采样点是合法采样点。

零件在拆卸路径的第二个采样点也被强制定义为合法采样点,这时因为,第一、二采样点的位置决定了零件可以自由移动的方向,即初始拆卸方向。

零件在拆卸路径的最后一个采样点,即零件被放置到工作台上的位置,也是合法采样点。

#### (2) 有效路径

零件的有效路径指的是由该零件的合法采样点依次连接而成的路径,且在该路径中的每一个位置,零件均不与其他零件发生碰撞。可以将零件的有效路径作为在用户指导的零件拆卸时记录的一种优化路径。

尽管设计者可以在三维工作空间中随意地移动零件,只要不发生碰撞;但是,零件在其拆卸过程的路径上应该具有一致性,即沿着某一运动方向具有一定的惯性。

## 3 典型虚拟装配系统功能介绍

目前世界很多国家都在致力于虚拟装配技术的研

究,而且开发出了一些系统。其中较为突出的有美国华盛顿州立大学和国家标准技术研究所 NIST 联合开发的虚拟装配设计环境 VADE 系统。开发 VADE 的目的是通过生成一个用于装配规划和评价的虚拟环境来探索在设计和制造中运用虚拟环境的技术可能性。

该系统以 SGI Onyx2 (6 个 processors, 2 个 Infinite Reality Pipes) 为平台,以 Flock of Birds、Cyberglove 和 VR4 头盔为虚拟现实交互设备。VADE 系统的结构和信息流如图 41.6-5 所示。

VADE 的主要功能特性为<sup>[23]</sup>:

1) 从 CAD 系统到 VR 的自动数据转化 VADE

自动将参数化 CAD 系统(如 Pro/E)中的产品装配树、零部件的几何形状传递到 VR 系统中。

2) 从 CAD 系统中捕捉装配意图并应用于虚拟环境 VADE 通过对 CAD 系统中装配约束的捕捉,实现对零件运动的引导与装配序列的生成。

3) 零件的交互动力学模拟 在物理模型基础上进行实体的碰撞检测,模拟用户、零件、装配工具以及装配环境之间的动力学作用。

4) 扫掠体积生成的轨迹编辑 VADE 允许用户记录、编辑零件的装配轨迹,然后在虚拟环境中生成于显示零件的扫掠体积。

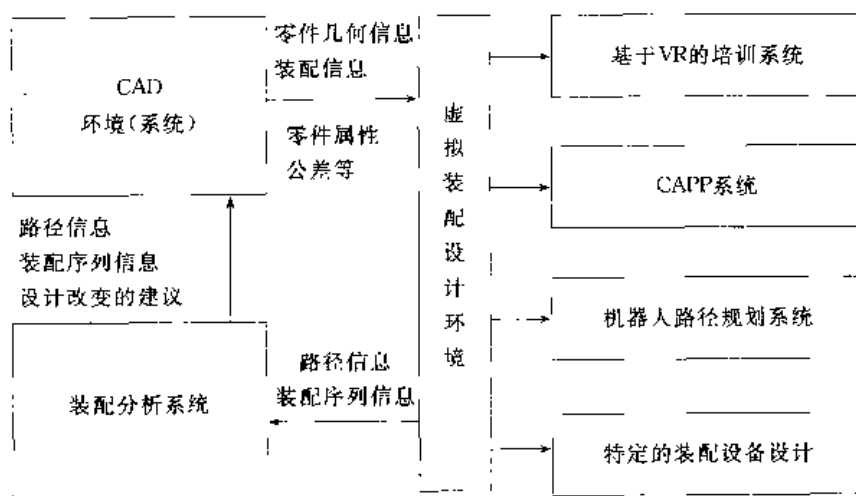


图 41.6-5 VADE 系统结构和信息流

5) 虚拟环境中对零件结构参数的修改 首先,VADE 将 CAD 系统中标识的零件模型的关键参数提取出来供用户在虚拟环境中修改,最后将修改后的零件重新传入 VADE。其中,参数传递与 CAD 系统对零件结构的参数修改均在后台进行,无需用户干涉。

6) 装配环境与零件初始位置的生成 整个装配环境可以在 CAD 系统中定义,同时用户可指定零部件的装配初始位置。

7) 双手装配与灵活操作 VADE 同时支持单手与双手操作。双手操作时,佩戴的手套设备很灵活,已有的算法支持能对其拿着的部件进行紧密操作,另一只手可用来抓住和操纵子装配的基础部件,使得其他零部件能够装配到它上面。

8) 支持虚拟装配工具 虚拟装配工具是装配环境的重要组成部分,VADE 提供了“手—工具”与“工具—零件”两种交互方法,并通过这两种方法的协同进行虚拟环境中零件运动的控制。

而在实际生产应用中,虚拟装配技术的优势已得到很大体现。目前已有虚拟装配技术应用于大型企业的成功实例,如美国波音飞机公司采用虚拟装配技术检查波音 777 飞机装配过程中零件间的干涉情况,并确保所有零部件对不同身材的装配工人来说都是可接触的;福特公司较早地将虚拟现实技术应用于装配仿真中,在 C3P 项目中,采用虚拟现实技术进行汽车的设计与装配,以确保产品的可装配性、易装配性和人机性能。

## 第7章 面向产品开发的工程应用

### 1 应用背景介绍

某型低地板城市客车是上海某公司的最新车型，现应用于上海市的主要道路交通。该型客车完全是虚拟设计的产物，2000年，上海市提出要提高城市交通车辆的整体水平，并在排放、乘坐舒适性、稳定性、适合上海路况要求等多项严格要求，国外的城市客车造价昂贵，且设计不完全符合中国人习惯。国内传统的客车设计方法不能够解决这些问题，要求在半年时间甚至更短时间就要设计出合格产品，并且出样车，在这些要求下，虚拟设计迎合了这种需求：

1) 在客车没有造出来之前，就可以在计算机中产生客车的模型，不需要制造实物模型，可以节约成本。

2) 修改客车样式，只需在计算机中更改，一处更改，其他立刻更新，极大缩短设计时间。

3) 客车的各种性能，通过计算机仿真可以随时仿真模拟，改变设计不合理。

4) 客车的舒适性，通过中国人体模型来验证乘坐舒适性。

### 2 虚拟客车车身开发系统的集成平台

在多家开发单位分布、异构的环境下，建立基于创

新设计的虚拟车身开发的集成平台，实现基于PDM的产品数据及开发过程的集成化管理，以支持虚拟客车车身的分布异构环境下的集成开发。

应用系统集成平台的体系结构分为控制层、应用层、对象管理层和支撑层，如图41.7-1。控制层完成对过程的监控与管理。应用层面向产品开发的不同应用领域，完成不同的任务，如CAD、CAE、CAM、DFA、各种专家系统等等，该层是数据的产生器和接受器，它通过应用工具接口封装在对象管理框架中，实现“即插即用”；在系统中应用层分成总布置设计、外形设计、结构设计、内饰设计、虚拟样车和工艺设计几个大模块（这些模块的内容将在后续章节中论述，除工艺设计模块外），启动应用层执行的是产品预研后得到的设计任务书，工艺设计完成后进入生产设计。对象管理层以EDS公司Metaphase的OMF（对象管理框架）为核心，实现对车身设计过程中所涉及到的对象进行定义和管理的目的。支撑层为上层提供支持，包括网络、操作系统和数据库以及分布、异构的计算机环境。

围绕汽车车身复杂产品的开发过程，进行分布异地的多单位合作，产生了极其庞大的数据并形成复杂的开发流程。下面将对集成平台开发中的数据仓库规划和流程管理进行论述。

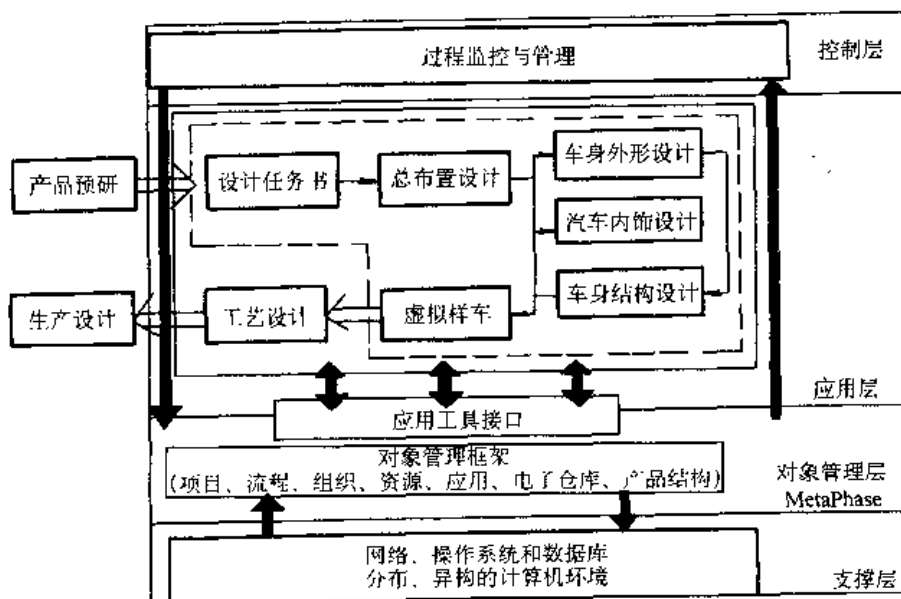


图 41.7-1 虚拟客车车身开发的应用系统体系结构

## 2.1 数据仓库规划

通过对数据仓库进行逻辑上分层,数量上分散,控制上分级,物理上分布的“四分式”规划,可以让用户对数据仓库的访问更加合理、高效和安全。

分层表现在将元数据(所谓元数据,是用来描述应用数据的数据对象,用户通过查询元数据获取与之相关联的应用数据)与应用数据分开存储。元数据库存放在服务器,应用数据仓库则单独存放于数据库系统之外,在 Metaphase 环境下,采用称为“文件系统”的存储空间来予以实现。文件系统是 Metaphase 系统用来存放应用数据文件的地方,一个文件系统对应操作系统中某一个指定的目录。一般来说,用户的文件在共享之前都放在个人工作区,个人工作区必须放在文件系统中,因此为了便于管理用户的个人数据,必须在每台主机上都至少设定一个文件系统。当然,过多的文件系统,也会给管理带来不便,因此除个别用于测试或特殊用途的主机外,每台主机上设置且仅设置一个文件系统。

分散表现在将用于共享的应用数据仓库离散成多个子仓库。将产品开发人员按其承担的角色分成多个用户组,每个用户组设有一个用于共享的子仓库。系统中共定义了 13 个主要用户,13 个对应角色。

分级则按照用户与数据仓库之间的访问权限关系,将数据仓库分成个人工作区、局部共享区和全局共享区几个级别。

### (1) 个人工作区

在 Metaphase 环境下,个人工作区称为 Work Location,一般情况下,每个设计、开发人员在其能够登录的主机上至少拥有一个个人工作区,用于存放私有数据。由于个人工作区是用来管理每个人自身的私有数据的,所以 Work location 的管理权完全由各个用户自己控制,用户可以根据自己的需要创建一个或者多个 Work location,在存放一个数据文件时指定个人工作区作为存放位置。

### (2) 局部共享区

在 Metaphase 环境下存放共享信息的空间称为 Vault,共享信息包括元数据信息和应用数据信息,对于元数据来说,其信息存放于数据库中,不需与文件系统发生关系;对于应用数据来说,其信息存放于文件系统中,因此需要指定具体的存放子目录,称为 Vault Location。与一个用户可以拥有多个个人工作区 Worklocation 一样,一个 Vault 可以拥有一个或多个 Vault Location。

### (3) 全局共享区

在系统中共设置了两个全局共享区,一是 AdminVault(对应的应用数据存储空间为 AdminVault),二是 ShareVault(对应的应用数据存储空间为 ShareVault),前者用于设计数据的归档,权限控制极为严格,对这一共享区的更新(包括添加和修改等)操作都由流程自动完成;后者用于存放设计过程中全局性的共享信息,如产品设计规范、项目进度计划等等,普通用户对该数据仓库拥有查看的权限。这种多级规划的整体结构见图 41-7-2。

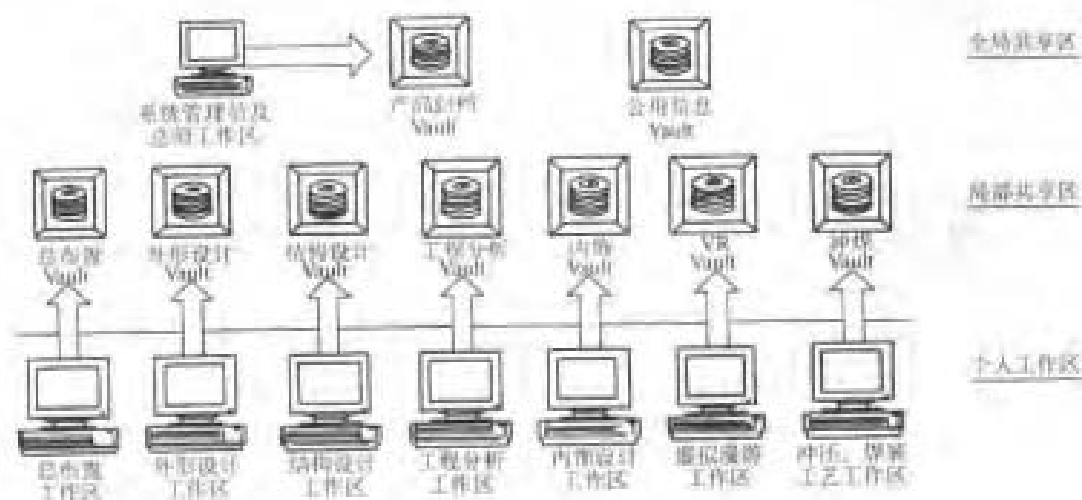


图 41-7-2 实施数据仓库规划

分布操作将分散操作形成的多个数据仓库按照访问其内容的设计人员所在地域特征的不同予以适当的分布,通过优化数据的传输路线来达到降低网络负载和提高访问效率的目的。分布操作的宗旨是让用户尽

可能从离自己近的数据仓库中找到所需要的数据。

## 2.2 流程管理

### (1) 流程分析与规划

流程管理的重要内容是在系统中建立与实际过程相符合的电子化设计流程,而建立电子化设计流程的前提是要对产品开发过程的各个环节进行信息需求分析,确定车身开发过程中各个模块及子模块之间的输入输出关系。

经过分解与细化并详细分析活动间的依赖关系,形成了贯穿客车车身开发过程流程,见图 41-7-3。图

中每个节点除了标明活动名称,还标明开发人员及其所属的合作单位,该流程在 Metaphase 中的表达形式见图 41-7-4,由于受篇幅所限,图中仅给出了一级流程。一级流程中的每一个节点都分别包含一个子流程,有些子流程下还有更下一级的流程,整个系统采用嵌套的方式逐层实现,子流程的具体表达细节此不赘述。

说明:

1. 流程可从实现、验证两个角度看;
2. 每个活动指明了相应的参与人员、格式、“单位、人性”;
3. 略去了与 Metaphase 系统相关的活动。

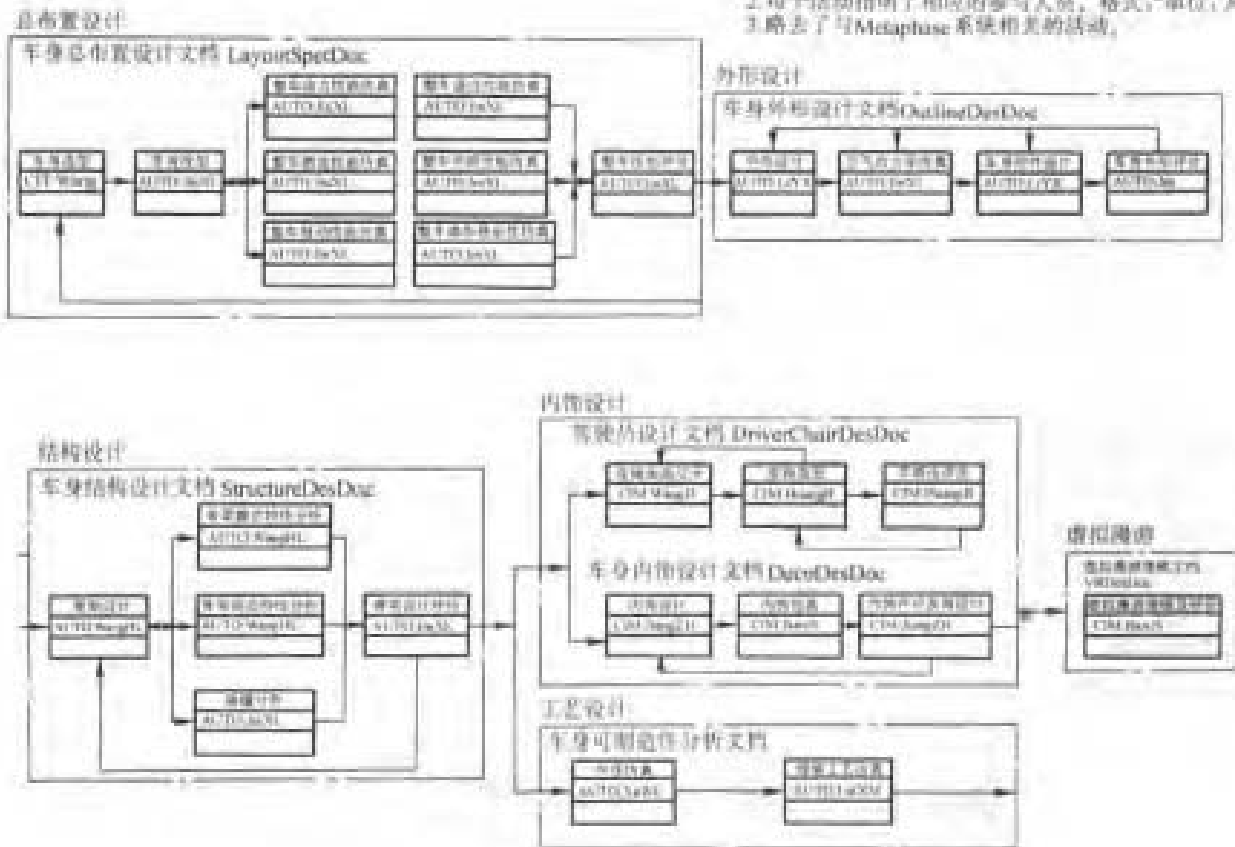


图 41-7-3 车身开发流程

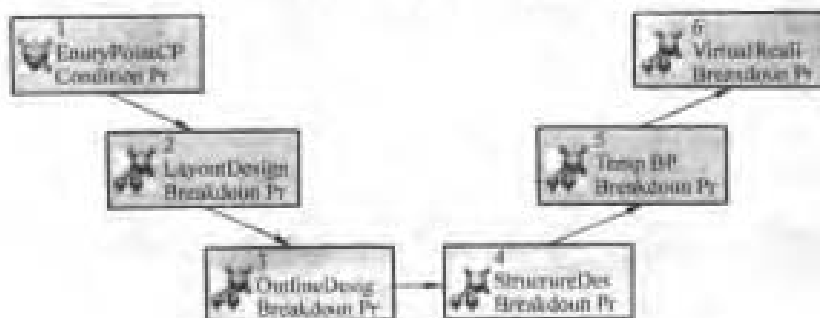


图 41-7-4 车身开发的一级流程 (在 Metaphase 中的表示)

(2) 多反馈机制及实现

在多数 PDM 产品提供的流程管理模块中,一个签审过程被拒绝后,只能将流程反馈到一个指定的过程,但在实际企业产品开发中,通常一个过程的反馈是

不确定的,不同条件下签审过程被否决后会反馈到不同的前序活动,由此可见目前 PDM 系统所提供的流程管理功能有着不完善的地方。

由于车身开发流程中涉及的活动较多,为了尽可

能真实地反映车身的设计过程,引入多点反馈机制,实现的途径是利用C语言以及Metaphase软件提供的Model语言等工具进行二次开发。

二次开发的基本思路是在Metaphase界面中的Signal菜单下增加一个名为Return to process的菜单项(见图41.7-5)。签审活动中用户单击该菜单项时,出现一个对话框,在对话框的列表框里显示出同一个流程中当前签审活动前的所有开发活动的名称(见图41.7-6)。这样用户就可以在这个列表框中看到他要反馈到的开发活动名称,从而可将流程反馈到该开发活动。

在图41.7-6中,右边对话框列表框里列出的是在当前过程前面已经流过的设计活动名称,当用户选中其中之一并单击OK按钮后,该活动名称便会显示到左边对话框中的文本框中,接着单击左边对话框中的OK按钮和图41.7-5中的Reject菜单项,便可以将流程反馈到选取的设计活动去。

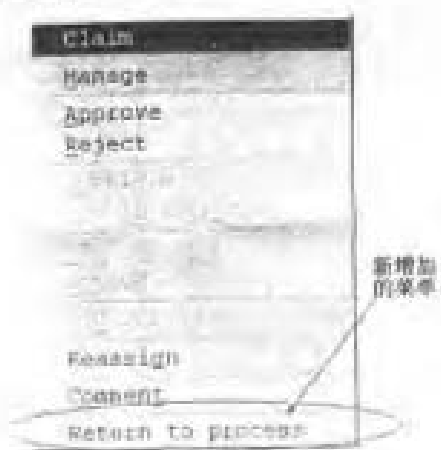


图 41.7-5 新增菜单项



图 41.7-6 反馈点选择

### 3 车身外形设计与性能分析

车身的外形设计流程如图41.7-7所示。



图 41.7-7 车身的外形设计流程

#### 3.1 客车车身曲面设计

客车车身是由几个曲面组合而成的,每个曲面的外形是设计人员在造型的基础上拟合而得到的,并通过比较模拟,使车身外廓线连接光滑,过渡平顺,使客车整体统一,比例协调,并具有动感,常规曲面或形体有两种:一曲线平移法;二曲线绕某个轴线旋转成形体。在客车车身造型中,经常使用的是圆锥曲线,对圆锥而言,它上一条封闭曲线,形状简单易于掌握;同一圆锥上任一点曲率相同,没有变化,在造型上显得呆板,不生动。对于椭圆、抛物线、双曲线而言,其上任意一点的曲率都是变化的,而且随着点的移动,变化逐渐减慢或加剧,呈一定的规律,在造型上显得生动,有“弹性”。椭圆是一个封闭结构的曲线,而抛物线和双曲线不是封闭曲线,因此,此处车身曲线以椭圆曲线为主,车身各曲面的结构曲线参数是由车身布置图的曲线上采点拟合而成,并将得到的数值经过适当调整,以便车身曲面间更协调。

客车车身前视图的视图曲线通常为一段圆弧或椭圆或三段圆弧相切,中心侧视图曲线一般为两段圆弧相切或椭圆与圆弧相切或椭圆曲线,前视图曲面的形成是以前视图曲线为截面线,前图的中心侧视图曲线为轨迹线



构造的,两条曲线的参数是通过在图上量取合适的点,经过拟合和调整得到的,两条曲线分别组合为两个椭圆曲线,以同样的方式,分别得到客车车身侧围、后围、顶盖曲面的构造曲线参数,几个曲面的构造曲线拟合为直线或椭圆,利用平移法或旋转法形成曲面,因此,利用车身曲面横截曲线或轨迹曲线的长短轴作为设计参数,通过调整这些设计参数就可以改变车身曲面的形状。

### 3.2 客车车身外形性能分析

对于客车车身在空气中运动时所受空气对车身作用力的问题,所研究的对象是不可压缩的黏性流体对车体的作用,当汽车速度超过 $60\text{km/h}$ 时,汽车的空气阻力占总阻力的一半以上,这时就必须研究其空气动力学特性,降低阻力系数,提高汽车的动力性和燃油经济性,同时高速行驶对空气动力稳定性也有较大的影响,提高汽车高速行驶稳定性的要求也使得对汽车的空气动力学研究提出要求。

车身造型空气动力学特性的检测主要依靠风洞实验完成:将选定的车身造型缩小比例缩成模型放入风洞实验场,通过观察空气流场烟雾或染料经过处理的模型表面的变化来进行相关性能的测试。但是,风洞实验存在着以下问题。

1) 由于模型与实物大小不一,实验结果与实际情况存在一定误差。

2) 为了与实际情况符合,就必须使用 $1:1$ 的车身模型,而大尺寸的风洞造价昂贵,结构复杂,只有军工、航空航天等特殊部门和机构才拥有且数量极少,汽车企业很难实现。

3) 每当车身造型设计需要修改,则必须重新制作实物模型,重新进行风洞实验。

在该低地板城市客车车身的设计中,利用计算流体动力学方法,通过沉浸可视化软件,在计算机上建立虚拟风洞,来评价车身外形的设计。

#### (1) 客车气流场分析的数据准备

在Pro/E软件中设计车体的外形,如图41-7-8所示,以几何模型为基础,不考虑外形的复杂细节,因车身纵向轮廓对称,进行CFD计算时只需要计算一半,尺寸大小: $11850\text{mm}\times 3250\text{mm}\times 2490\text{mm}$ 。将一半车体放在一个足够大的空间中,删去掉本身的车体,对剩下的空间区域划分网格,采用三角形网格,约包括401010个单元,如图41-7-9所示,客车在中速行驶时,表面的流体属性是粘流,表面速度为零,定义好边界条件,迭代方法等,利用CFD软件对其进行计算,计算结果如图41-7-10所示,从图中可知,流场数据可视化效果较弱,考虑借助沉浸可视化的虚拟环境,进行计算

数据的直观显示。

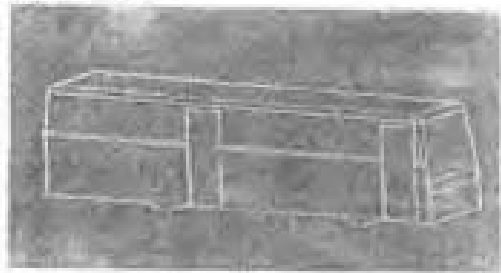


图 41-7-8 客车三维几何模型



图 41-7-9 网格模

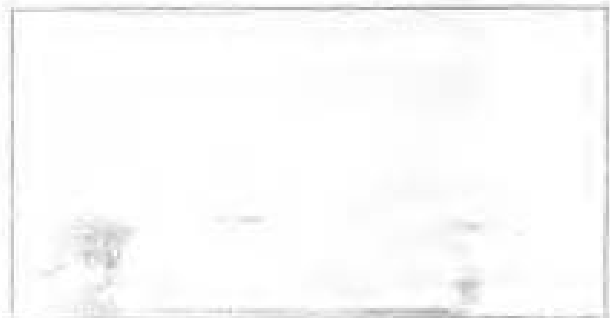


图 41-7-10 客车风洞 CFD 计算

#### (2) 客车气流场的沉浸可视化分析

因为计算后的数据量非常庞大,一个工况的计算量,包括三个方向上的速度场、压力场总共数据量 $131\text{M}$ ,如果包括导出量,数据量要扩大 $5\sim 10$ 倍。用显式过滤方法,首先过滤去掉网格模型,然后搜索所有关键点,构建流线拓扑结构,然后计算流线的几何特征和物体属性,判断是否存在涡,只保留涡的拓扑结构,便于分析。

如图41-7-11为渲染后具有真实感模型的客车虚拟风洞可视化,从图41-7-10中可看出在汽车后面有两个垂直方向的涡,计算结果符合Volvo公司的风洞试验,结合真实感的几何模型,在虚拟现实环境中,通过可视化不同速度下的流场,观察流线的形态、涡位置和强度,判断客车外型设计是否符合气动力学,是否需要修改外形,如通过修改接近角和离去角以及客车后

部形状,减少尾部的涡大小,增加客车在运行过程中的平稳性和阻力,以获得最佳的气动特性的客车外形。图 41.7.12a 中客车离去角  $6^\circ$ , 图 41.7.12b 中客车离去角

$8.5^\circ$ , 可以看出在离去角为  $8.5^\circ$  时, 客车尾部涡最小, 从而减阻系数最小, 在实际的生产中, 就是采用离去角为  $8.5^\circ$  的外形。

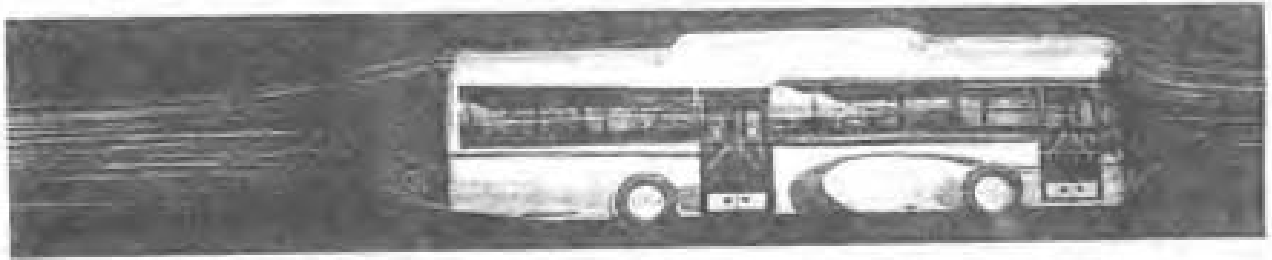


图 41.7.11 客车风阻设计可视化

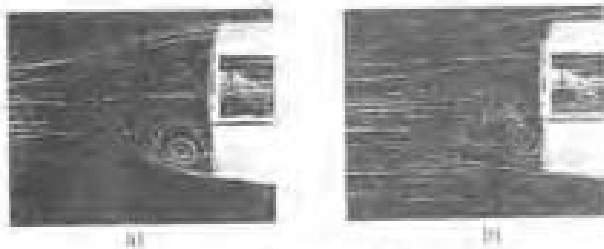


图 41.7.12 不同离去角尾流涡流状态对比

#### 4 结构设计及性能分析

车身的结构设计内容如图 41.7.13 所示。

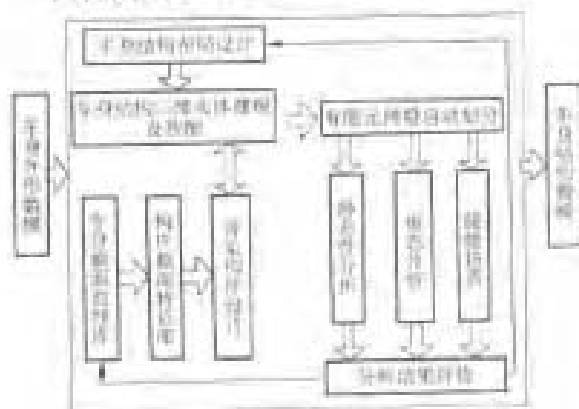


图 41.7.13 车身的结构设计

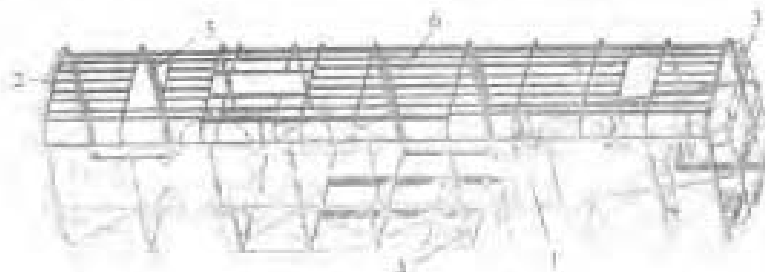


图 41.7.14 低地板城市客车车身骨架结构

1—轮架 2—前围骨架 3—后围骨架 4—左侧围骨架 5—右侧围骨架 6—顶盖骨架

利用 Pro/Engineer 提供的布局设计 (Layout) 功能, 可以在客车车身设计的初始阶段定义一个装配体

#### 4.1 基于三维模型的车身结构设计及装配

低地板城市客车车身主要采用骨架蒙皮式结构形式。客车骨架结构是由抗扭刚度较高的矩形钢管构件焊接形成的空间框架。通常由底架、前、后围骨架, 左、右侧围骨架和顶盖骨架六部分组成。

利用基于特征的成组技术, 使用零件主模型思想, 建立车身骨架关键构件库。充分利用客车车身骨架构件截面特征和车身截面曲线特征较为规则单一的特点, 将构件截面特征分为矩形、槽形、圆形、工字形、L 形 (角钢) 及薄板等类型, 将车身截面曲线特征分为椭圆、圆弧、直线段等类型, 应用成组技术, 通过建立和使用零件主模型, 快速生成特征相关、尺寸参数不同的零件族, 并初步建立了车身骨架关键零件库, 有利于实现结构设计中装配建模的自动化和结构的快速再设计。

装配建模所要表达的主要信息是各零部件之间的相对位置关系 (几何约束), 对于像车身这样的大型复杂装配体, 存在着装配层次、装配顺序和设计参数等诸多关键要素, 通过合理确定装配层次、装配顺序, 充分利用设计参数和几何约束的表达, 能够方便快捷地修改和重构装配模型, 从而实现结构设计过程的并行化。

的基本约束要求, 基本结构和全局参数及参数之间的顺序依赖关系, 其零件部件可以是一个概念上的方块

图形或参数草图,然后建立参数、尺寸之间的关系和零部件自动定位所需的全局基准。

#### 4.2 基于三维模型的车身结构静强度分析

结合城市客车实际运行中出现的多种典型工况,如直线匀速行驶,路面高低不平出现单轮瞬间悬空,紧急制动及急速转弯等工况,研究相应载荷及边界约束条件施加的实现方法,分析车身结构强度、刚度。

对低地板城市客车车身结构,应用有限元法进行静力特性和动态特性分析,还分别用梁、板壳单元计算其截面应力和变形。梁单元仅能总体反映车身的应力分布情况,对于复杂的连接结构,通过使用与实际结构更为接近的板壳或实体单元能取得较高的计算精度。

通常对于城市客车,仅计算满载状态下的弯曲和扭转工况,但由于低地板城市客车采用空气弹簧悬架,在紧急制动和急速转弯时,纵向力和横向力必须通过推力杆传递,车身承载状况复杂,有必要校核其结构强度。结合城市客车实际运行中出现的多种典型工况,如直线匀速行驶,路面高低不平出现单轮瞬间悬空,紧急制动及急速转弯等工况,提出了相应载荷及边界约束条件施加的实现方法,进而对整车的静动态特性进行了分析与比较,结果与实际使用情况基本相符。

##### (1) 车身骨架结构离散模型

通过IGES接口,可将Pro/E三维CAD设计软件环境下建立的车身骨架结构模型转换进ANSYS有限元分析软件环境中。

由于车身骨架结构不对称,故有限元计算采用整体结构的模式。由于车身骨架是由不同厚度的矩形管、槽钢、角钢和平板组焊而成,它们既可承受平板内的载荷,又承受垂直于板平面的载荷,所以在建模时采用了弯曲板单元进行离散。初步计算时,考虑到计算精度和计算机容量与运算速度的协调,故单元划分的不至于过细,计11690个单元,13573个结点。结构离散模型见图41.7-15。

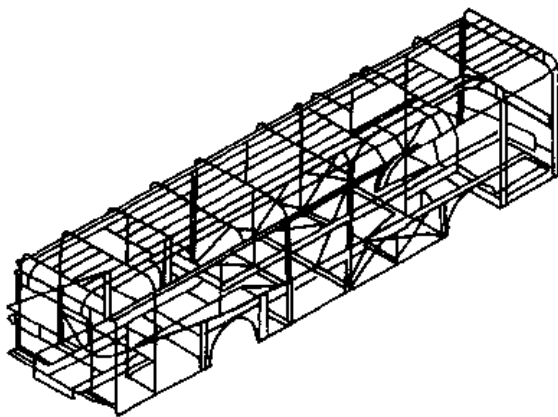


图 41.7-15 车身骨架结构离散模型

##### (2) 空气弹簧悬架结构离散模型

为准确模拟各种计算工况下的边界条件,对后车架的空气弹簧悬挂(图41.7-16)采用如图41.7-17所示的等效计算模型,共计5个结点4个单元。单元1、4是线性弹簧单元,采用combin14模拟;单元2、3是大刚度梁单元,承载时变形极小,不致影响弹簧单元的计算。

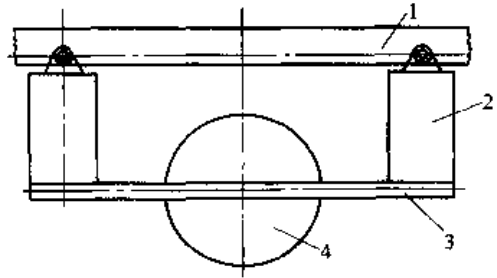


图 41.7-16 后车架空气弹簧悬挂结构示意图

1—车架 2—空气弹簧 3—扁担梁 4—车轮

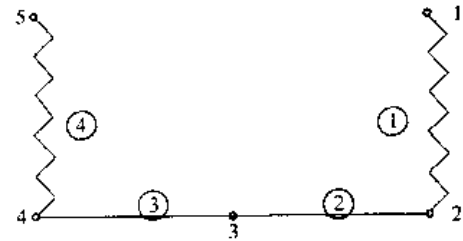


图 41.7-17 后车架空气弹簧悬挂等效计算模型

1~5—结点号,①~④—单元号

表 41.7-1 材料参数

名称	弹性模量 /(N· mm <sup>-2</sup> )	泊松 比	密度 /(kg· mm <sup>-3</sup> )	强度极限 /(N· mm <sup>-2</sup> )	屈服强度 /(N· mm <sup>-2</sup> )
09SiV	2.1×10 <sup>5</sup>	0.3	7.85×10 <sup>-6</sup>	550	330
A3	2.1×10 <sup>5</sup>	0.3	7.85×10 <sup>-6</sup>	380	230

##### (3) 材料

底架主要采用09SiV低合金结构钢,其他车身分片骨架均采用A3普通碳素钢,具体参数见表41.7-1。

##### (4) 计算工况

“汽车产品定型可靠性行驶试验规程”规定:样车必须以一定车速,在各种道路上行驶一定里程。典型工况主要包括高速道路、一般道路、弯道行驶,行驶时会出现静弯曲、扭转、紧急制动和急速转弯等四种工况。大多数文献多集中讨论前两种工况,而对后两种出现频次较高(特别是城市客车)的工况未加分析。计算分析时,应对可能出现的各种工况均予考虑,才可能确定车身结构强度和刚度是否满足要求,以及进一步进行优化设计。图41.7-18所示为静弯曲工况下的车身结构变形分布图。

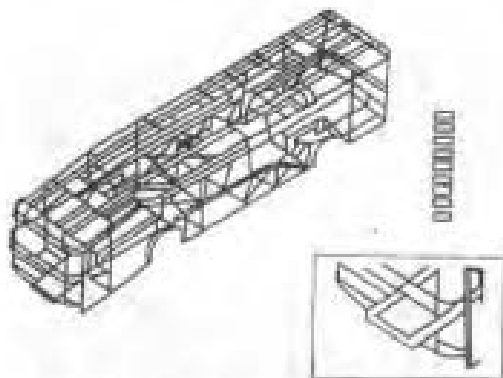


图 41-7-18 静载荷工况下轿车车身结构应力分布图

### 4.3 基于三维模型的车身结构模态分析

客车在运行过程中,由于路面的不平整,会受到一定的动载荷的作用。当所受动载荷的频率与车身结构的某一固有频率接近时,可能引起结构共振,从而产生很高的动应力,造成强度破坏或产生不允许的大变形,破坏汽车的性能。因此必须对客车的结构进行模态分析。

通过模态分析可以确定设计中的客车的激励特性(固有频率和振型)。模态分析也是进行更详细的动力学分析的基础。例如谱响应分析,瞬态动力学分析,谱分析。模态分析所用的离散模型与结构静态分析时所用的模型相同。

模态提取方法:典型的无阻尼模态分析求解的基本方程是经典的特征值问题:

$$K\Phi = \omega^2 M\Phi$$

式中:  $K$ ——刚度矩阵;

$\Phi$ ——第  $i$  阶模态的振型向量(特征向量);

$\omega_i$ ——第  $i$  阶模态的固有频率( $\omega_i^2$  是特征值);

$M$ ——质量矩阵。

有多种数值方法可用于求解上面的方程。其中 Block Lanczos 法具有运算速度快,输入参数少,特征值、特征向量求解精度高的特点,特别适用于模型中包含形状较差的实体和壳单元的大模型。表 41-7-2 为低地板客车车身骨架前五阶模态计算结果。

表 41-7-2 车身骨架前五阶模态计算结果

阶数	1	2	3	4	5
固有频率/Hz	3.42	5.061	6.359	7.411	8.285
振型	车身前 后摇摆	车身后 后摇摆	车身后 侧扭转	扭转	车身后 侧弯曲

### 4.4 大客车整车的碰撞安全性数值模拟

汽车碰撞过程是一个涉及大位移,大应变,大转动的复杂非线性问题。在分析求解这种问题时,除了考虑几何非线性,还要考虑材料非线性和状态非线性。其

中,几何非线性包含两层含义,一是位移与应变之间的关系是非线性的,在小变形中忽略了二次导数项,简化为线性关系,二是变形过程包含有刚体转动,在小变形中也忽略了刚体转动的影响。如要精确地求解性大变形的非线性问题,这些因素就不能忽略,同时也大大增加了问题的复杂性。

有限元方法作为工程领域中复杂问题数值解析与模拟的重要工具,各发达国家都投入了大量的资金和人力对其进行研究和应用开发,其商用软件也日益增多。然而所有的有限元分析商用软件均可分为三种类型:动力显式软件(如 DYN3D, PAM-STAMP, OPTRIS),静力隐式软件(如 ITASSD, ROBUST)和静力显式软件(如 MARC, ABAQUS)。

建立客车的有限元模型,包括壳单元、梁单元,进行大客车的碰撞分析。共有节点 13312 个,壳单元 15027 个,梁单元 30 个,客车整车的有限元模型如图 41-7-19 所示。图 41-7-20 为在某个时刻大客车整车变形图。

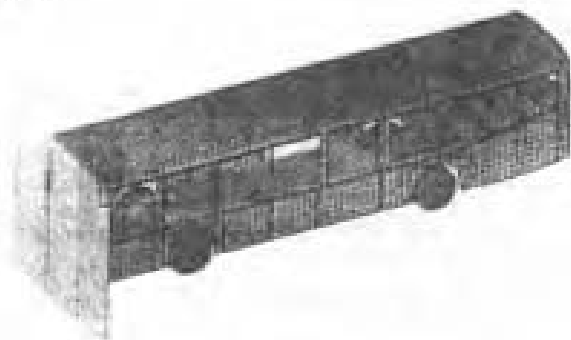


图 41-7-19 大客车的碰撞模拟有限元模型

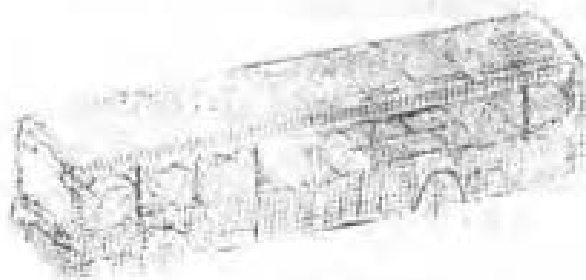


图 41-7-20 大客车整车变形图

图 41-7-21 为大客车驾驶员座椅固定点的加速度曲线,可以看出,计算结果比通常的轿车车身的加速度高的多,这是因为大客车的车身结构受到冲击后,没有轿车的前部碰撞吸能区,不能及时吸收碰撞动能。从整个客车车身结构也可以看出,车身的骨架结构为刚性连接,没有专门的碰撞性能结构。一般国内大客车的结构设计主要考虑强度和刚度,而汽车的结构耐撞性能

还没有引起重视。

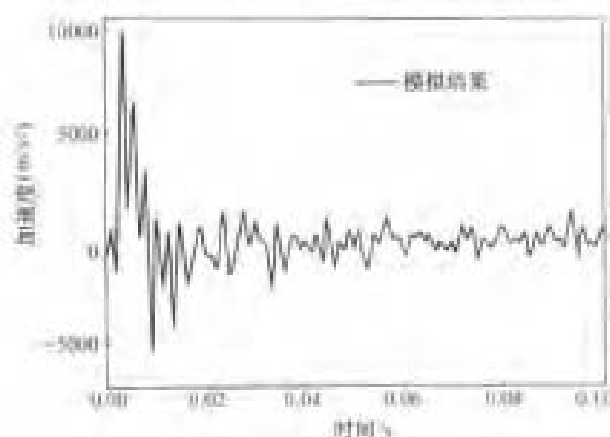


图 41.7-21 大客车驾驶员座椅固定点的加速度 (客车纵轴方向)

## 5 内饰设计与性能分析

汽车的内饰很多,如座椅、仪表、控制器等,本节仅讨论座椅的设计分析问题。在利用虚拟产品开发技术进行客车内饰开发过程中,客车驾驶员座椅的虚拟设计和分析是一个重要环节。大量调研表明,我国汽车座椅方面的设计主要是靠仿造和经验,厂家基本上没有做过相应的定量分析,这主要和座椅定量分析对技术要求、费时费力有关。随着人体舒适性得到广泛关注,座椅的静态和动态性能成为汽车性能的重要评估内容。

汽车座椅的乘坐性能包括静态舒适性和动态舒适性<sup>[1]</sup>,前者主要与尺寸参数、表面质量、调节特性等有关;后者则主要与座椅系统的动态参数,如刚度、阻尼系数及振动特性有关。

### 5.1 司机座椅分析模型的建立

产品开发的基础是基于几何的多视图实体模型。20世纪80年代以来,知识工程和计算机仿真加强了基于几何的设计开发方法的应用。座椅虚拟设计与分析也是基于以下模型。

#### (1) 基于几何的实体模型

在座椅设计理论与标准的基础上,设计座椅系统的几何尺寸、结构、材料等方面的要素,并通过 Pro/Engineer 软件建立驾驶员座椅的三维几何模型。图 41.7-22 表示座椅三维几何模型。

#### (2) 动力学分析模型

按照各个分析模型视图的要求加入激励和约束,得到结构分析模型,如运动学动力学模型等。由于座椅主要的考察指标是静态舒适性和动态舒适性,因此主要分析集中在几何模型、结构分析模型和动力学模型。如图 41.7-23 为 ADAMS 中的动力学模型。

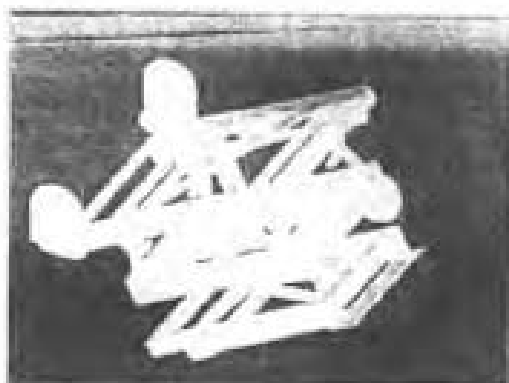


图 41.7-22 座椅系统三维几何模型



图 41.7-23 为 ADAMS 中的动力学模型

### 5.2 座椅的动态舒适性分析

座椅系统动态舒适性主要与乘坐的阻尼系数和刚度系数以及座椅基架的阻尼与刚度有关,因此,对座椅系统进行动力学分析时,主要考虑上述几个参数的选取对座椅动态舒适性的影响,并通过分析结果选取合适的设计参数。如图 41.7-24 为不同参数下人体驱动加速度时域和频域谱。

通过对座椅系统的动力学仿真试验,得出结论,根据用户要求,设计的驾驶员座椅的乘坐的刚度系数最佳取值范围为 12~14N/mm,阻尼系数最佳取值范围为 0.2~0.4Ns/mm。

根据仿真分析的结果,可以确定如下参数,座垫刚度系数 13,阻尼系数为 0.3Ns/mm,座椅悬架系统刚度系数 100N/mm,阻尼器的阻尼为 40Ns/mm,输入实测的汽车传入座椅系统激励时域谱,得出的加速度加权均方根值为 0.7554m/s²,小于客车驾驶员座椅的动态舒适性指标 1.12m/s² 允许值,这个结果说明设计的座椅符合动态舒适性的要求。

如上所述,在座椅的虚拟设计和分析中,可以在虚拟的开发条件下测试座椅的静态和动态性能,从而可以在不经过建造物理样机的情况下,确定座椅的几何信息和非几何信息,并对其性能有了全面的了解,使得虚拟样机能够代替或部分代替物理样机。

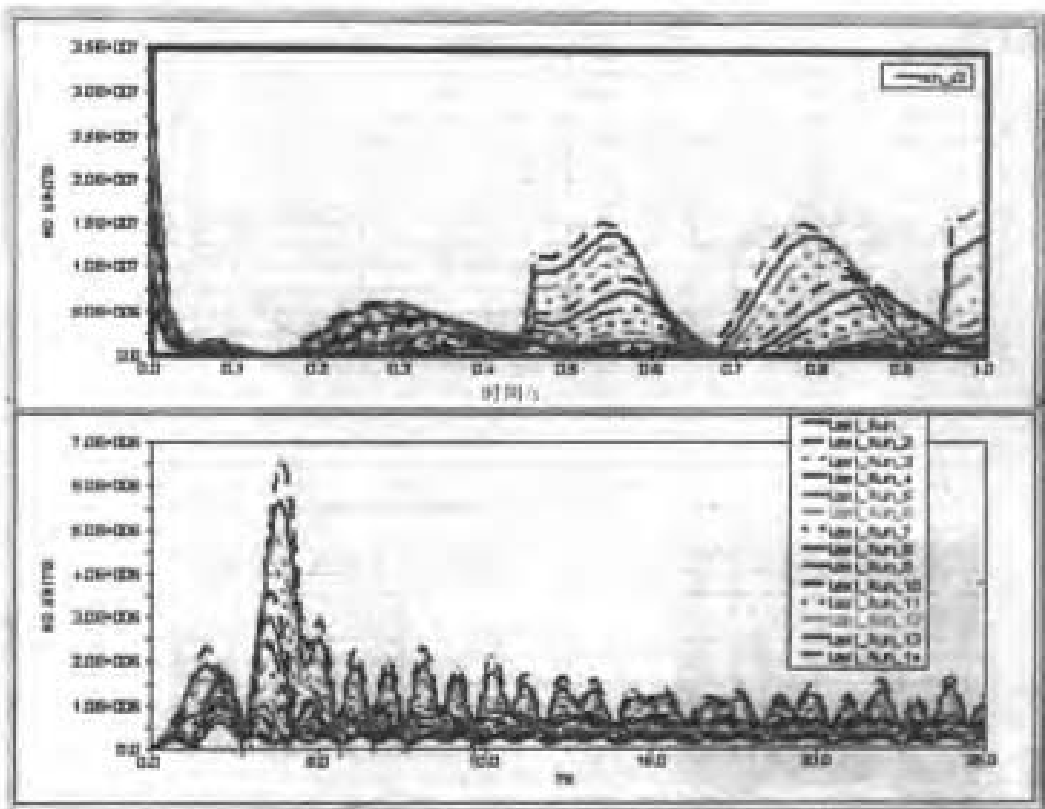


图 41.7-24 不同参数下人体振动加速度时域和频域谱

### 6 虚拟客车样机和人机工程分析

在客车的设计过程中，需建立精细的内饰模型以研究并评估人的因素和人体工程学问题。制作这些实物原型昂贵、耗时且难于修改，这样必然带来设计周期长，用虚拟原型取代实物模型，在设计早期就考虑产品与其他各个部件的关系，在虚拟环境中模拟真实的运行环境，是当今世界先进制造技术发展中的一个趋势。

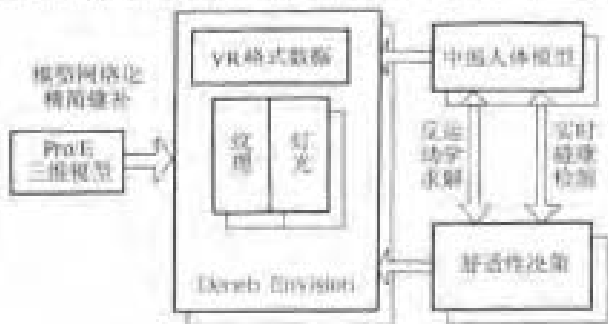


图 41.7-25 客车虚拟样机和人机工程分析

结合超低地板的客车内饰设计，研究客车内饰的空间舒适性等问题，用虚拟原型来代替实物原型。如图 41.7-25 所示，在 Deneb/Envision 环境中，进行虚拟客车样机的相关操作：客车内部布局和装饰、仪表、控制器的可见性分析；可及性分析；碰撞检测；人体空间舒

适性分析；艺术造型和外表渲染等。

#### 6.1 虚拟客车样机的建立

##### (1) 实体数据的转换

构造虚拟原型过程中的难点之一是几何模型的精简。虚拟现实中的物体不需要实体模型也不能是实体模型，是由点、线、面片构成的，而三维软件的实体模型往往包括大量自由表面（如 NURBS、B 样条曲面），需要大量的多边形来逼近模型，虚拟环境的沉浸程度取决于模型的真实程度，但是过多的细节程度（LOD）又会影响系统的实时性，在沉浸的漫游过程中，必须保证系统的实时渲染响应速度达到 20~30 帧/s，必须通过精简模型以达到系统的实时性要求，图 41.7-26 是模型精简的对照图。这里采用了边界折叠算法，开发模型精简软件，可以高效地解决模型精简中的多边形重组、缝合等，同时维护了模型的拓扑结构。经过同样数据转换的还包括司机座椅、仪表盘、客车车身模型等。

##### (2) 渲染技术的应用

计算机图形技术是让模型栩栩如生地体现在用户眼前，其中主要的技术有纹理技术、光照技术、模型精简，让虚拟产品更加自然、逼真地体现在用户眼前，对于设计以及用户感受模型的演化非常重要。图 41.7-27 所示为一个座椅渲染后的场景。

##### (3) 人体模型的建立

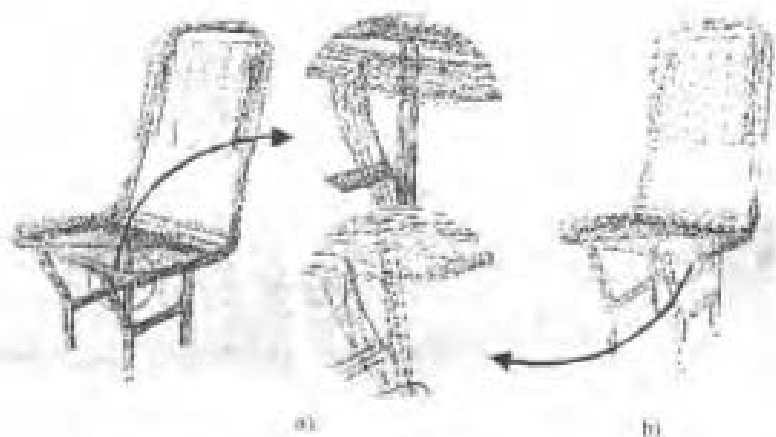


图 41-7-26

a) 未精简的乘客座椅 (2348 个面片) b) 精简后的乘客座椅 (1348 个面片)

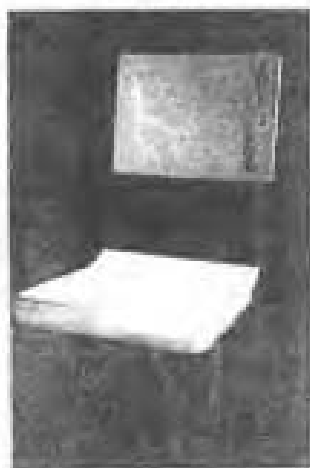


图 41-7-27 虚拟后的乘客座

建立数字化的人体几何模型,引入虚拟现实系统中。在 Deneb 的 Ergo 模型的基础上,结合中国人口普查结果,建立符合中国人体数字模型,建立头部、手部、躯干、腿部的运动链,模拟中国人的行走、站立等。

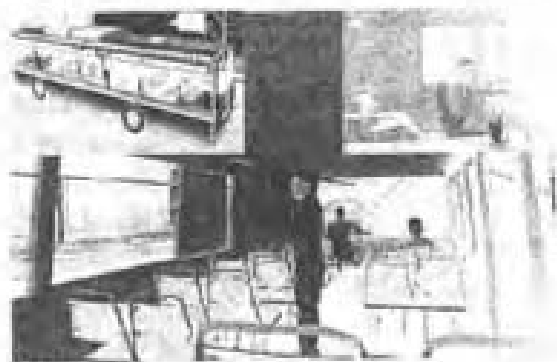


图 41-7-28 虚拟客车原型

通过上述三方面的处理,进行客车内饰布置、纹理渲染等建立虚拟客车原型(由于数据量所限,虚拟客车样机中没有包含客车的所有零部件),如下图 41-7-28 所示。

## 6.2 人机工程分析

根据人机工程学的知识,进行客车内饰的舒适性分析。按照全息产品建模的原理,对于产品座椅,环境就是人体模型,研究的对象是座椅本身属性,人体模型与座椅的交互结果反映了座椅的设计是否合理。是否符合人机工程。人的任何运动是相互协调的,进行人体舒适性分析,必须基于反运动学计算,实现对人行行为的跟踪,从而评价舒适性。借助 Ergo 模型可以实现反运动学计算。

对所建立的客车原型,定义虚拟漫游路径,进行人体动作仿真,佩带立体眼镜,实现虚拟漫游,进行如下的人体舒适性分析:

### (1) 乘坐舒适性分析

驾驶员在行车过程中,精力高度集中,非常容易疲劳,因此驾驶员座椅的乘坐舒适性对于安全行车非常重要。在虚拟环境中,让人体模型“坐”上司机座椅,调节座椅的空间位置,建立司机驾车模型,感受座椅的舒适性。根据国家标准对驾驶员座椅的舒适性定义,如图 41-7-29 所示的驾驶员的驾驶位置两种方案中,图 41-7-29b 为不舒适情况。通过改变驾驶员乘坐的角度,以满足座椅舒适性要求,最终确定座椅的位置、座椅靠背的倾斜角度等。

乘坐空间指乘客座椅间的空间,如果太小,乘客感到挤,如果空间太大又影响到整个客车的乘员数目。这里乘客空间的确定,通过国家对乘客座椅的排列标准以及通过人体模型确定的舒适程度给定,在布置座椅时,将上述条件作为座椅的碰撞检测条件,使乘客空间的设计更加合理。

### (2) 可及性分析

驾驶过程中,驾驶员需要操纵许多仪表和操纵杆,通过人体模型来仿真驾驶员在行车过程中对不可仪表

的操纵情况,用以验证仪表和操纵杆的可及性,使得驾驶员非常方便地操纵各种仪表和操纵杆。评价的标准是不影响驾驶员操纵的坐姿,方便操纵各种仪表。

### (3) 视野分析

视野是驾驶员操作时,不需改变操作姿势对道路及周围环境观察的可见范围。驾驶员的视野是保证汽车操纵方便和行驶安全的重要条件之一。驾驶员的视野必须满足正常驾驶坐姿下,随时可清楚地看到按道路法规所设置的一切指示和警告标志。

确定视野取决于座椅的位置、高度及坐垫和靠背的倾角,车窗的尺寸、形状和布置,立柱的结构等。对视野的评定是以驾驶员眼睛的“椭圆点”为起点划出汽车上各部件到前方的距离,具体规则可参照国家标准。传统的视野设计非常复杂,需要大量的二维投影计算,而在虚拟现实环境中,通过引入人体模型,建立视野模型去“看”,驾驶员看到的就是真正的视野,比起传统的视野设计需要大量的计算要方便很多。同时可以实时地修改各种影响视野的参数,获得非常理想的视野。

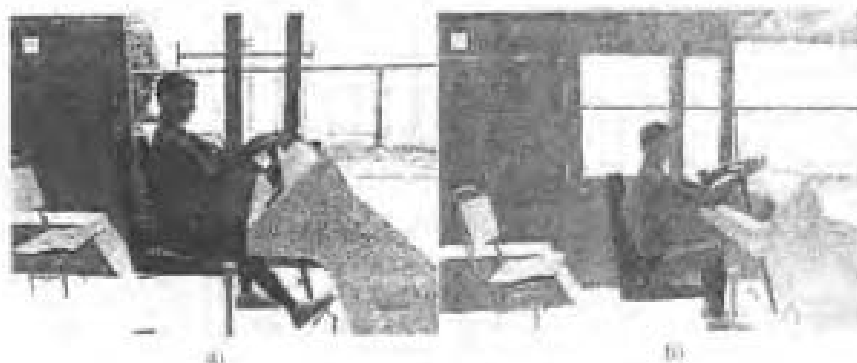


图 41.7-29 驾驶员的乘坐位置

### 参 考 文 献

- 1 严真琪,范秀敏等.虚拟制造的理论、技术基础与实践.上海:上海交通大学出版社,2003:26~30,114~115
- 2 汪成为,高文,王行仁.灵境(虚拟现实)技术的理论、实现及应用.北京:清华大学出版社,南宁:广西科学技术出版社,1996
- 3 John Vince.Virtual Reality Systems. Addison-Wesley Publishing Company, 1997
- 4 曾建超,俞志和.虚拟现实的技术及其应用.北京:清华大学出版社,1996
- 5 McCormick B. H., DeFanti T. A., Brown M. D. Visualization in Scientific computing. Computer Graphics, 1987, 21 (6), 5~7
- 6 唐泽圣.三维数据场可视化.北京:清华大学出版社,1999
- 7 石教英,蔡文立.科学计算可视化算法与系统.北京:科学出版社,1996
- 8 Grigore Bardea, Philippe Couffé.Virtual Reality Technology. John Wiley&Sons, Inc., 1994
- 9 Steve Bryson.The distributed virtual wind-tunnel. Supercomputing, 1992, 18 (9)
- 10 Bejura M., Fuchs H., Uchiuchi R. Merging virtual objects with the real world; seeing ultrasound imagery within the patient. Computer Graphics Proceedings, Annual Conference Series (26), 1992, 203~210
- 11 Cruz-Neira C., Sardin D. J., DeFanti T. A., Kenyon R. V., Hart J. C. The CAVE; audio visual experience automatic virtual Environment. Communications of the ACM, 1993, (35): 67~72
- 12 李翠兰,史义前等.计算机辅助制造技术综述.天水师範学院学报,2001,21 (5): 53~54
- 13 Tsung-Pin Yeh, Judy M. Vance. Combining NASTRAN sensitivity methods and virtual reality to facilitate interactive design. Finite Elements in Analysis and Design, 1997, (26)
- 14 Michael J. Ryken, Judy M. Vance. Applying virtual reality techniques to the interactive stress analysis of a tractor lift arm. Finite Elements in Analysis and Design, 2000, (35):
- 15 Hiroyuki Tao, Katsumi Tadamura. Multi-functional visualization system for structure Analysis. IEEE, 1997
- 16 Liverani A., Kuester F., Hamann B. Towards interactive finite element analysis of shell structures in virtual reality. IEEE, 1999
- 17 Nagraju Rangaraju, Michel Terk. Framework for immersive visualization of building analysis data. IEEE, 2001



- 18 Mike Connell, Odd Tullberg, A framework for immersive FEM visualization using transparent object communication in a distributed network environment. *Advances in Engineering Software*, 2002, (33)
- 19 Bao J S, Jin Y., Gu M.Q, Yan J.Q., Ma D.Z., Immersive virtual product development. *Journal of Materials Processing Technology*, 2002 (129):
- 20 Fabio Arciniegas. *XML Developer's Guide*. Mc Graw Hill, Inc., 2001; 9~15
- 21 NCSA HDF GROUP. HDF5 HOME PAGE, <http://hdf.ncsa.uiuc.edu/HDF5>, 2004
- 22 Mark P, *MSC/Patran User's Guide (2nd Edition)*. USA; The MacNeal-Schwendler Corporation, 2000
- 23 Lorensen W E, Cline H E. Marching Cubes: A high resolution 3D surface construction algorithm. *computer Graphics*, 1987, 21 (4): 163~169
- 24 Gallagher R S, Nagtegaal J C, Karlsson H, et al. An efficient 3-D visualization technique for finite element models and other coarse volumes. *Computer Graphics*, 1989, 23 (3), 185~194
- 25 William J Schroeder. *The VTK User's Guide*. USA; Kitware, Inc., 2002
- 26 *OpenGL Performer Programmer's Guide*. USA; Silicon Graphics, Inc., 1997
- 27 Paul J R, *vtkActor ToPF*, <http://brighton.ncsa.uiuc.edu/prajlich/vtkActorToPF/>, 2003
- 28 Matthias Kalle Dalheimer. *Programming with Qt*. O'REILLY, Inc., 2002

# 第 42 篇 快速响应变型设计

主 编 钟廷修  
编写人 钟廷修 杨洪柏  
孙 莉 赵继云

# 第 1 章 绪 论

## 1 快速响应工程及其含义

近十几年来,世界经济和技术的发展势头良好。由于社会生产力高度发达,市场上产品日趋丰富,消费者的选择余地日益扩大,所以对产品提出越来越高的要求,使用户需求呈现出差异而多变,这迫使企业跟随用户,采用多品种、小批量的生产方式,按定单组织生产。同时,面对瞬息万变的市场环境,更要求企业不断地迅速开发新产品,以吸引顾客,这样才能保证企业在竞争环境中立于不败之地。可见,在这种时代背景下,市场竞争的焦点就转移到速度上来,即凡能“领先一步”,快速提供更高性能/价格比产品的企业,将具有更强的综合竞争力。因此,我们提倡实施“快速响应工程”,以适应市场环境的变化和用户需求的转移。

实施“快速响应工程”,可以使企业具备从捕捉产品信息到及时投放市场的全面快速响应能力。其主要内容包括以下几个模块:

- 建立快速捕捉市场动态需求信息的决策机制;
- 实现产品的快速响应设计;
- 追求新产品的快速试制定型;
- 推行快速响应制造的生产体系。

为了提高快速响应能力,企业首先应能迅速捕捉复杂多变的动态市场信息,并及时作出正确的预测和决策,以决定新产品的功能特征和上市时间。明确了新产品的开发项目以后,实现快速响应设计就成为重要的一环。工程设计的重要性是不言而喻的。据统计,工程设计的费用虽然只占产品最终成本的一小部分(不到10%),但往往决定了它的80%以上的制造成本,而且还决定性地影响产品的性能和交货期。现代机械产品由于用户的要求越来越高,产品结构日益复杂,科技含量愈来愈高,所以使得产品的开发周期日趋延长。如何解决好产品市场寿命缩短和新产品开发周期延长的尖锐矛盾,已经成为决定企业成败兴衰的生死攸关问题。产品开发周期包括设计、试制、试验和修改等一系列环节,除了设计以外的后几个环节可以统称为试制定型阶段。为了实现快速设计,人们提出了CE(并行工程)、DFX(面向制造、装配、检验、质量、服务等的设计)、CSCW(计算机协同工作支持环境)和功能分解组合的设计思想,这将引起对现代设计方法和CAD发展的新探索。这些将在“并行设计”一文中有专门的论述。在此基础上,本章提出用变型设计实现快速响应的

策略。这种设计策略特别强调对企业产品信息的标准化、规范化重组,通过对企业现有成熟产品的变型再设计,使企业的宝贵信息资源得到尽可能多的重用,从而达到以快交货、高质量、低成本和重环保的快速响应竞争策略去赢得市场的目的。

加快产品的试制、试验和定型,以快速形成生产力,需要尽量利用制造自动化的各种新技术,如FMS、RP(Rapid Prototyping,快速成型)和VM(Virtual Manufacturing,虚拟制造)。快速成型堪称20年来制造技术最重大的进展之一,其特点是能以最快的速度将CAD模型转换为产品原型或直接制造零件,从而使产品开发可以进行快速测试、评价和改进,以完成设计定型,或快速形成精密铸件和模具等的批量生产能力。虚拟制造是先进制造技术的先导,它通过计算机仿真和多媒体技术全面模拟现实制造系统中的物流、信息流、能量流和资金流,可以做到在产品制出之前,就能由虚拟环境形成虚样品(Soft Prototype),以替代传统的实样品(Hard Prototype),进行试验和评价,从而大大缩短产品的开发周期。

推行产品的快速响应制造,必然导致企业从组织形式到技术路线的一系列变革。首先,在企业内部,应改变传统的以注重规模和成本为基础建立起来的生产管理系统和组织形式,按照快速响应制造的战略思想,探索一套全新的组织生产方式,例如将生产部门从以功能为基础的工序组合改变为以产品为对象的加工单元,并且尽量采用各种先进制造技术手段,等等。其次,从面向全局的视野出发,以产品为纽带,以效益为中心,不分企业内外、地域差异,实行动态联盟,有效地组织产品的设计、制造和营销。企业在确定产品目标后,可以只先进行总体设计,即功能设计、方案设计和经济分析,然后通过公共信息网络,寻找最佳的零部件供应商和制造商,进行跨地区、跨行业的合作,实行生产资源的优化组合,并由承包商按照快速响应的原则进行具体设计,即结构设计、详细设计和工艺设计,组织产品的快速响应制造,以保证产品及时上市,经由遍布各地的营销网络迅速抢占市场。

由此可知,实施快速响应工程,至少有如下优点:及时交货,争先抢占市场,因缩短制造时间和消除各种浪费而降低制造成本,因加快产品更新换代而向市场大量提供质量更高、价格更低的高性能产品。其结果就大大提高了企业的综合竞争力和持续发展的生命力。

## 2 快速响应设计是实施快速响应工程的重要一环

对企业生产活动的大量调查表明,产品设计制约着包括技术准备(试制、工艺、工装、采购)在内的产品生产周期。为了实现快速响应市场需求,进行快速响应产品设计是重要一环。实现快速响应设计的关键是有效开发和利用各种产品信息资源。利用现代化计算机技术、信息技术和通讯技术迅猛发展所提供的对信息资源的高度存储、传播和加工能力,人们主要采取三项基本策略,以达到对产品设计需求的快速响应。这三项策略,一是重用产品信息和知识资源,进行变型设计;二是虚拟设计制造,利用数字技术加快设计过程;三是远程协同,分布设计。概括起来,就是信息资源化,产品数字化,设计网络化。本篇着重介绍对产品信息和知识资源的重复利用。企业在长期的生产活动中积累和蕴藏了大量的极其宝贵的产品资料(图样、文件、数据、经验、标准、规范等),对这些信息和知识进行重组使之资源化,再通过对成熟产品的变型设计,就足以支持能快速响应市场需求的产品设计。其重点是:1)重组企业工作流程,将开拓未来市场的“慢节奏”新型产品开发与面对当前需求的“快节奏”变型产品设计区别开来,用变型产品设计快速响应市场需求;2)重组企业产品信息,使企业在设计开发领域中积累的丰富技术资料,转变为有用和好用的宝贵信息和知识资源,以期尽量利用已有的产品信息(尤其是机电产品中的成熟零部件)设计新产品,即用尽可能少的新设计零部件,配制成尽可能多的变型产品;3)利用信息技术和设计理论的最新成就,以期通过产品结构的重组来支持快速变型设计。简而言之,就是业务过程重组、信息资源重组和产品结构重组。本篇将详细阐述这项策略思想的体系结构、关键技术及其实现方法。

## 3 用变型设计实现快速响应的方法

机械产品的设计,通常可分为新颖性/新型设计和适应性/变型设计两大类。新型设计基于全新的工作原理,采用基本新颖的结构方案,设计出创新的机械产品。这类设计需要经过从确定设计要求到完成全部技术文件的整个设计工作进程。变型设计基于已有的工作原理,采用基本不变的结构方案,只按功能需求对具体结构进行局部调整,以产生适应性的变型产品。这类设计因工作原理不变,仅对局部结构或零件进行更新设计。无论是创新设计还是变型设计,均体现了设计人员的创造性思维,都能提供新的、具有社会价值的产品。

常用的创造性思维方法和创造性技法有:

### (1)直觉思维法

该方法充分利用直觉思维和灵感思维来得到问题的新解法或新的发现。其中又分智爆法、集智小组法和专家预测法。

### (2)推理思维法

推理思维法也是一种重要的创造性技法。它的特点有二:一是在逻辑推理、启发思考和强制联想的基础上扩展思路;二是在分析的基础上推理,即经过分析,把一个复杂问题细分为一些较为简单的小问题,再通过分析每个问题的各种影响因素,有针对性、有步骤地进行推理,寻求解法。较为典型的推理思维法有:奥斯本提问法、阿诺德提问法、参量提问法、反向探求法和系统搜索法。

### (3)联想创造法

通过启发、类比、联想、综合等技法,创造出新的设想,以解决问题。常用的有:相似类比法、抽象类比法、仿生法、借用法。

### (4)组合创新法

利用事物间的内在联系,用已有的知识和现有成果进行新的组合,从而产生新的方案。合理的组合也是一种创造。常用的组合创新法有:组合法和综摄法。

从企业的产品信息资源在设计活动中的可重用程度考察,新型设计涉及的可重用信息资源较少,它的实现过程需要耗用可观的企业资源(人力、物力、财力和时间等),以验证与完善产品结构;变型设计在原有产品的基础上,按市场需求进行结构重组,它的实现过程可以最大限度地重用企业已有的成熟产品资源,具有很强的灵活性和适应性。可见,变型设计可以在原有产品的基础上迅速推出能满足不同用户需求的新产品,占领市场,增强企业的竞争能力。

变型产品具有以下特点:

1)灵活 根据市场需求,灵活地推出多种相应产品,要善于“变”。

2)迅速 为适应市场竞争的需要,推出产品要快。

3)低廉 在保证功能和质量的前提下成本要低,或确切地说性能价格比要低。

在变型产品的设计中应贯彻标准化、通用化和系列化的原则,以提高产品质量,降低成本,得到多品种多规格产品。

标准化指使用要求相同的零件、产品或工程,按照统一的标准进行设计。通用化是指在同一类型、不同规格或不同类型的产品中,提高部分零件或部件彼此相互通用的程度。系列化是指产品根据生产和使用的技术要求,经过技术和经济分析,适当加以归并和简化,将产品的主要参数和性能指标按一定的规律进行分档,合理地安排产品的品种规格以形成系列。标准化、

通用化、系列化均属于标准化的工作范畴。标准化工作是国家的一项重要经济政策。它对保证产品质量、缩短新产品研制和生产周期、便于使用维修、降低成本等各方面都有重大意义。标准化水平是衡量一个国家技术管理水平的尺度,也是现代化的一个重要标志。

变型产品设计的基本方法有相似设计、模块化设计、产品族设计和基于实例设计等。

相似设计是相似理论在设计中的应用。系统相似,尺寸与性能参数皆成一定比例关系的系列产品为相似系列产品。相似系列产品不是按单个产品设计的,而是首先选定某一中档的产品为基型,对它进行最佳方案的设计,确定其材料、参数和尺寸,然后再按系列设计原理,即通过相似理论求出系列中其他产品的参数和尺寸。前者称为基型产品,后者称为扩展型产品。

目前,产品的系列化设计成为广泛应用的设计方法之一。它的优点是:

1) 系列产品的不同规格仅仅是基于一种规格变化而形成的,这就大大节省了产品的开发周期和成本,提高了产品的性能可靠性。

2) 系列产品在满足用户需求的前提下,遵循适当的参数变化规律可以提高不同规格产品的生产批量,从而使产品质量稳定、成本下降,这对生产企业和用户都是有利的。

3) 对生产和销售企业来说,系列产品便于库存管理;对用户来说,系列产品的使用规定和方法相同,方便了使用。

在进行系列化设计时,首先要进行市场调查和用户访问,并在此基础上进行基型产品设计。基型产品的选择首先要注意它的代表性,它应是所有形式中最通用的,基型产品的典型结构应能很方便地用到同系列其他产品中去。总之,基型产品设计是发展产品系列的基础,其设计过程应遵循科学的设计方法,并注意其扩展的可能性和方便性,注意提高其标准化、通用化的程度,注意尽可能提高其设计质量。有了基型产品后,便可确定相似种类和系列各参数或尺寸间的公比相似比。一般生产者希望产品规格少、批量大,即要求公比大;但使用者却希望规格多,能较好地符合使用要求,即要求公比小。因而确定相似比时需兼顾生产和使用的要求,并尽量选用标准公比和标准数列。在此基础上,便可确定系列中各扩展型的参数和尺寸,进而确定系列产品结构和尺寸。

模块化设计是近年来发达国家普遍采用的一种先进设计方法。它的核心思想是将系统按功能分解为若干模块,通过模块的不同组合,可以得到不同品种、不同规格的产品。从50年代欧美一些国家提出这一设计方法以来,它已扩展到很多行业,并与CAD技术、成

组加工技术、柔性加工技术等先进技术密切联系起来,应用到实际产品的设计与制造中。在机械产品中,所谓模块就是一组具有同一功能和结合要素,但性能和结构不同,却能互换的单元。

模块化设计是将产品上同一功能的单元设计成具有不同性能、可以互换的模块,选用不同模块,即可组成不同类型、不同规格的产品。模块化设计的原则是力求以少数模块组成尽可能多的产品,并在满足用户要求的基础上使产品精度高、性能稳定、结构简单、成本低廉。显然,为了保证模块的互换,必须提高其标准化、通用化、规格化的程度。模块化设计应首先用于系列产品设计中。

采用模块化设计产品有以下优点:

1) 产品更新换代较快 新产品的发展常是局部的改进,若将先进技术引入相应模块,比较容易实现。这就加快了产品的更新换代。

2) 可以缩短设计和制造周期 用户提出要求后,只需更换部分模块,或设计、制造个别模块即可获得所需产品。这样设计和制造周期就大大缩短了。

3) 可以降低成本 模块化后,同一模块可用于多种产品,增大了该模块的生产批量,便于采用先进工艺、成组技术等,还可缩短设计时间,从而降低产品成本,提高产品质量。

4) 维修方便 必要时可只更换模块。

5) 模块化设计对产品的功能划分及模块设计进行了精心研究,这就保证了它的性能,使产品性能稳定可靠。

当然,模块化设计也有其缺点:由于考虑模块的互换性和适应性,常使系统比较复杂,结构外形不够协调,各部分功能配合不是最佳等。

由此可见,模块化设计用于生产批量较小的系列产品是特别有效的。

在进行模块化设计时,应以用户的需要为根据,通过市场调查及技术经济可行性分析,确定产品的系列型谱。如何合理划分模块,是模块化产品设计中的关键问题。模块种类少,通用化程度高,加工批量大,对降低成本较有利。但从满足产品的各种功能和性能,提高整个模块化系统的柔性考虑,又希望模块种类多些。设计时必须从产品系统的整体出发,对功能、性能和成本各方面进行全面分析,才能合理确定模块的种类和数量。划分模块的出发点是功能分析。根据产品的总功能分解为分功能、功能元,再求解相应的功能载体——功能模块,然后再具体化为加工单元——生产模块。模块化产品的设计中还要考虑模块如何组合、组合产品的种类,以求用较少类型模块组合成更多不同功能和性能的变型产品。值得注意的是,先进的模块系统不但可采

用计算机辅助设计,而且可用计算机进行管理,通过计算机辅助管理能更好地体现模块化设计的优越性。

产品族是一类具有相同功能和结构形状,针对一定范围内用户需求的一系列产品的集合。产品族系列设计能够根据预定的产品结构,即根据预定义的产品零部件集及它们之间的相互约束关系,按照用户的个性需求,只需通过一个常规的产品配置过程,就可迅速完成用户定制产品的设计。基于实例设计根据用户的个性需求,先从已往的产品实例库中提取功能相似的实例,然后对提取的实例进行结构和参数分析,如能满足当前设计的要求,则可作为目标方案完成设计,如果

只能部分满足设计要求,则应对其进行修改,直至满足所有设计要求为止。

快速响应变型设计的目的是加快产品的上市时间(交货期),手段是企业信息资源的合理定义和高效处理。因此,凡能有效利用企业宝贵信息资源的设计方法,都能实现快速响应。从这个原则出发,除了上述的系列化、模块化和产品族设计与方法之外,本篇将着重进行阐述的基于实例的变型设计方法,也能实现快速响应。而且,这种具有一定智能化程度的设计方法,更能体现变型设计的通用化、过程化和产品个性化特点。

## 第 2 章 快速响应变型设计的体系结构

产品设计是企业生产活动的一个组成部分,它受企业的生产规模、活动模式和产品复杂程度等多种因素的影响。从考察企业的生产活动中可以看出,企业中往往是新型产品开发设计和老产品变型设计两者并存,以适应市场发展的不同要求。

在机械企业中,新型设计毕竟稀少,大量的总是变型设计。事实上,绝大多数人员的确也是经常使用变型设计去响应用户需要。但是,传统的、自发的变型设计并不能自动保证产品的快交货、高质量和低成本,有时甚至适得其反。主要表现在企业生产零件的多样性和缺乏可重用性,并导致新零件数目的失控。许多新的零件被源源不断地设计出来,紧接而来的还需要动用过多的人力、物力、财力和时间,去支持后继的工艺、工装、试制和试验等一系列技术准备过程。引起这种失控现象的原因在于企业产品信息的无序状态和设计者的随意性,使大量的产品数据无法成为能有效支持快速变型设计的宝贵信息资源。因此,深入系统地研究一种能科学地重用企业的产品信息资源,有效支持快速响应设计的变型设计理论和方法及其实现手段,就很重要。

### 1 变型设计的过程

在传统的设计理论中,典型的设计过程是功能空间(Function Space)、行为空间(Behavior Space)和结构空间(Structural Space)之间的映射过程。其中,功能空间是设计需要描述的空间,行为空间是对设计方案的描述,而结构空间则是对具体属性的描述。这是一个逐步完善、反复迭代的过程,因为产品是复杂的,设计者的知识和经验是有限的,所以在三个空间之间的映射几乎不可能是一次性的,总是必须经过多次反复才能完成。图 42.2-1 表示传统的设计过程模型。图中,设计信息流动与设计过程步骤有机地结合起来,不仅反映出设计过程的顺序,也描述了其中的数据流动关系。

传统的设计过程模型总是从零开始进行产品设计,即从初始的功能设计到最后的详细设计,尽管可能需要多次反复,但总体上是从无到有(从抽象概念到具体形象)、自顶向下的,也就是 Top-Down 式。这比较适合于新型设计。而变型设计是从已有产品的原型出发,然后根据实际的功能需求进行局部改动,其设计的重

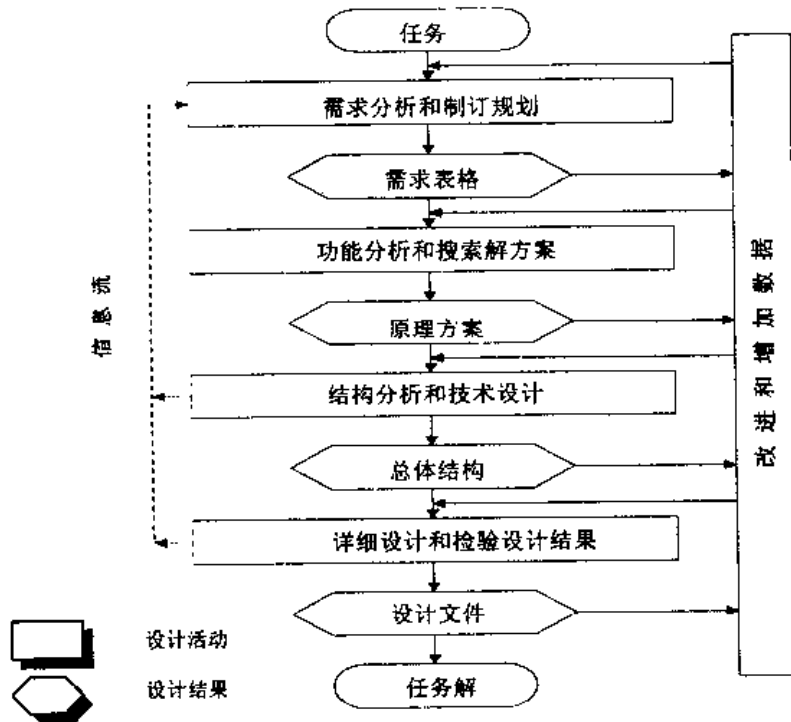


图 42.2-1 传统的设计过程模型

心在于搜索恰当的原型实例,然后通过分析比较,进行多层次的设计改动,因此其设计过程的信息流动和数据关系也与新型产品的设计过程有所区别,图 42.2-2 描述一般的变型设计过程。图中,变型设计一般从功能模型出发搜索设计原型,功能模型是分层次的,也就是总功能模型通常由许多分功能模块组成,因此设计原型也是分层次的,每一个设计原型对应一个或多个功能。设计师对设计原型的组成和结构进行选择、调整和

组合,有时直接从设计数据库中选择零部件,再通过结构重组使之符合设计要求,最终设计出新的变型产品。可见,对于变型设计产品来说,其设计过程模型往往是 Top-Down 式和 Down-Top 式的相结合,它在现有的产品资源的基础上分解产品结构,其功能空间、行为空间和结构空间都充满了空间元素——原型实例。所以,变型设计过程实际上是一个建立空间之间映射关系的过程。

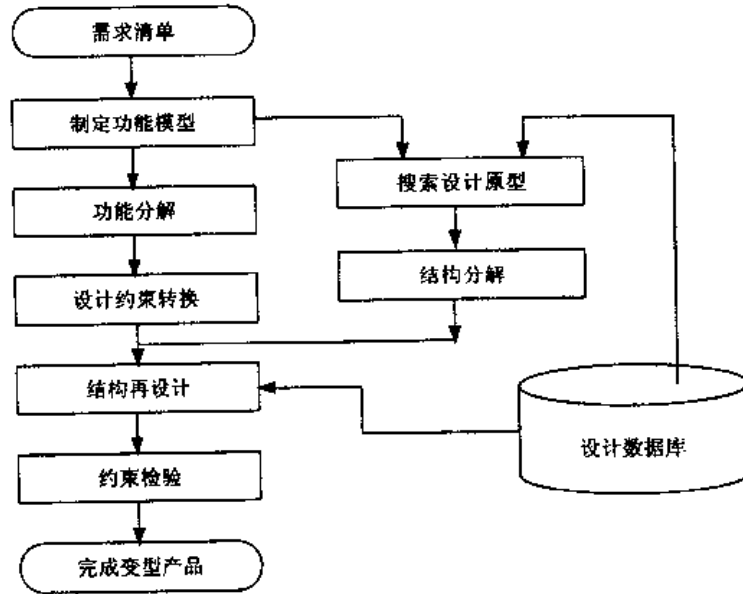


图 42.2-2 变型设计过程

## 2 变型设计的层次结构

上述在变型设计过程中,设计师往往将设计问题向已有的方案靠拢,再利用现有的知识来解决新的问题。对于比较复杂的设计问题,还采用“问题归约”的方法,即将复杂的问题分解为相对比较简单的多个子问题进行处理,这种分解过程一直可以持续下去,直到所有的子问题都得到解决,也就是整个设计问题的完成。可见,变型设计过程具有如下特点:

- 设计方案的形成由抽象向具体的方向发展,其过程是动态的,具有反复性;
- 设计对象的演变按总体—局部—总体的顺序发展;
- 设计目标参数相互关联;
- 设计对象的各部分属性部分耦合;
- 设计对象的分解具有层次性、结构性。

根据传统设计理论,整个设计过程划分为功能、行为和结构三个空间,产品的设计就是这三个空间的参数映射过程。实际上这样的映射是在不同的层次之间进行,因此变型设计系统的层次结构正是反映变型设计的渐进过程。上节中已详细描述了变型设计过程,说

明在设计产品的过程中所包含的信息并不是同一抽象层次的,如方案设计描述的是系统的原理方案,具有相当高的抽象层次,例如柴油机的连杆组件的原理解是曲轴连杆机构,对连杆组件的设计主要从曲轴连杆的运动参数出发。相反的,对连杆零件的设计则属于底层的详细结构设计,涉及的问题是零件的材料、几何形状、精度和加工特征等。其抽象程度比较低,所考虑的都是现实物理问题。变型设计系统要集成各过程的抽象信息,必须建立能处理不同抽象层次的体系结构,按抽象程度分为方案设计、结构设计、详细设计三个层次。

### 方案设计层

方案设计层主要处理设计任务描述、原理解设计、方案设计、原型确定。这些过程以功能为源头,以结构为目标,建立连接关系。四个处理过程间的数据流如图 42.2-3 所示。

方案设计层有三个基本功能:任务描述,功能结构关系描述,实例检索,以整个产品作为考虑对象,从产品级的角度分析对象关系,寻求合适的设计原型方案。任务描述是将设计任务要求转化为设计参数。这些参数必须建立在核心的关系型产品模型的基础上,是受



企业资源约束的。功能结构关系描述是用设计参数描述产品的功能,在此基础上寻求合适的原理解。原理解表达一定的功能结构,负担这些功能的物理结构体构成了方案集。每一类方案集都包含一定数量的原型设

计库。实例检索则包括实例的组织、存储和查询。实例是变型设计的基础,一个好的实例是变型设计成功的基石。从总体上分析,方案设计层是构筑设计对象的功能结构,将功能结构分解和综合,通过实例分析寻求

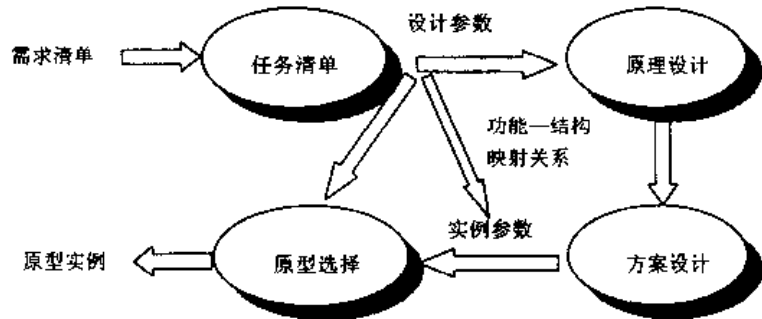


图 42.2-3 方案设计层数据流

合适的几何结构原型。

**结构设计层**

结构设计层的抽象层次要低一些,主要处理过程是结构变体分析、结构变体设计,其特点是结构—功能—结构。主要功能是结构功能描述、功能综合、结

构变体实现。结构设计层的数据操作是抽象、映射、具体化,而且操作对象是结构、功能。结构设计层是变型设计关键的一层,主要确定变型结构体,实现结构方案设计,确定变型产品的结构解,往往和装配模型有密切的联系,图 42.2-4 表示结构设计层的数据流。

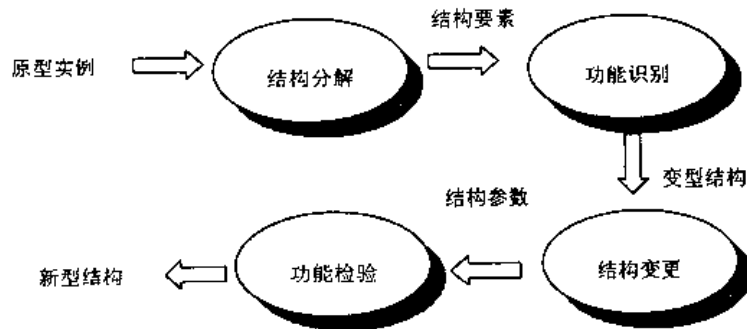


图 42.2-4 结构设计层数据流

结构功能描述是在原型结构分解的基础上,完整地描述整个系统每一部分的功能和结构的关系,产生功能—结构映射关系体。这种描述是基于关系型产品模型,整个系统的描述也是围绕产品内在的关系,功能综合根据关系和设计需求,断开部分需变动的结构,针对这些结构建立新的功能需求。结构变体实现根据这些功能需求,在产品资源库中搜索合适的资源,或者根据产品资源的规则建立新的产品资源,再由功能综合检验这些新的产品资源,看是否符合企业产品资源的规范和设计需求。

**详细设计层**

详细设计层是对新的产品资源和变动产品的具体参数进行确定,主要功能是资源参数确定、资源结构检验、资源特征检验。详细设计是对产品对象的具体属性

值的确定,从产品对象类中产生一个新的产品对象。通过在产品资源库中检索新产品资源的规则、参数范围、参数影响等,设计工程师选择或产生产品属性数据。每修改一个数据都进行特征和结构检验,避免发生干涉,如柴油机机体颜色采用绿色,对机体盖也必须采用绿色,另外如曲轴轴径改变,相应的支撑和轴承的相关尺寸也都应改变。

**3 变型设计的系统框架**

变型设计的系统框架如图 42.2-5 所示。整个变型产品设计系统主要包括三个部分:产品资源管理模块、产品设计管理模块、产品设计检验模块,支持部分包括数据库、通信接口。

**产品资源管理模块**

前已所述,快速变型设计的目的是快速开发出合适的产品,所以变型设计系统需要大量的可重复使用的产品资源,包括标准化的零件库、规范化的设计知识等。产品资源管理模块包括产品资源分类机制、产品资源定义机制、产品资源搜索机制。

产品资源分类机制中,产品结构树作为一种通用的分类方法,以产品为分类主码,如在某柴油机生产企业中,可按柴油机种类划分企业资源,如 135 柴油机、114 柴油机、180 柴油机等。在 135 柴油机中又根据产品对象的不同而划分成不同的资源类,如 G6135-000 型、G6135-001Z 型、7135-000 型等。在每一产品对象下以部件、零件、特征单元等分类组织产品资源,同时,在不同产品资源下有许多零部件、特征单元是相同的,可采用指针方式进行连接。产品结构树作为产品结构的一个直观表达,在结构中反映产品资源情况甚为形象,而且与现有管理方法相似,容易为设计师所接受,但是这样的管理方式存在数据冗余,而且当指针连接过于复杂时,数据管理容易混乱,有时甚至发生指针循环指向的情况。本系统主要采用另一种分类方法,按功能和使用环境分类,即将产品资源划分成若干使用功能对象类,如曲轴类、齿轮传动类、飞轮类等。在每一类下分为几个小类,如齿轮传动下可分为圆柱齿轮传动、圆锥齿轮传动等这样分,直到最后的具体的某一个实例对象。这样分类数据管理方便,而且层次鲜明,在设计过程中设计师可以按需求查询快速检索到合适的产品对象,但是如此分类也有结构不直观的缺点。因此本系统采用两种方法混用的方式,用功能方式储存产品资源,而用结构树建立结构目录。

一个物理的产品构成(零件、部件或产品)是以一定的信息形式表达出来的。这种形式可以是 2D 的二维图,也可以是 3D 的实体模型,还可以由文本来描述。在变型设计阶段,它们的管理信息除了涉及这些模型化的表达外,还将有命名、识别号、项目号、结构配置、BOM 表、设计者(及日期)、审批者(及日期)、更改标记(及责任者日期)等。这些信息,一方面是管理活动的需求,同时,也是其自身完备定义的需求。因此,将这些信息加以抽象、归类,就得到了产品资源的定义。

作为可重用的产品资源,完整的、独立的和唯一性的定义是必须的。在重用实例对象时,首先碰到的问题是检索。显然,每一个实例对象有如此多的属性定义,而每一个属性定义可以有很多的物理文档来表达,没有合适的检索策略的支持,就难以有效地检索需要的资源。每一产品资源都有其唯一的标识号,但识别号仅仅是机器内部的标识,不具有传统的隶属制下产品资源的定义(如图号),因此,直接使用识别号不仅仅是低效率,而且实际上也是行不通的。重用产品资源,一

方面,是指对一个产品资源对象类的实例对象的直接重用;另一方面,是指对它的属性定义的重用。因此,重用产品资源的检索过程也就是关于它的属性定义值的检索过程。系统建立产品资源对象及其属性字的检索目录,其中包括数据字典,它建立工程常用的术语与计算机化属性值的关系,通过检索目录可以快速检索到对象类的属性定义。产品资源的事物特性表中每一行记录是一个产品资源实例对象的属性值的描述,从表中直接可检索到它的事物特性记录、它的分类记录、它的视图定义、它的创建项目、它的重用记录、它的设计者等属性的值。

#### 产品设计管理模块

产品设计管理模块规划设计任务,分解并分配设计任务给不同的设计人员,并协调他们之间的关系,解决发生的冲突,还将设计结果存入数据库和知识库。

根据设计过程的不同阶段,产品设计管理模块还分成概念建模、装配建模和详细建模三个子系统。在概念设计过程中,设计师根据用户的需求,利用企业的设计知识和个人经验完成概念设计,形成产品的概念设计模型。概念设计的结果包括:功能结构层次关系图,功能映射图,设计参数网络等。由它们形成概念设计模型,这是下层装配建模的基础。装配建模是概念建模的继续,它受概念设计模型的约束,同时加以拓展。装配设计的结果包括:装配层次关系图、装配结构关系图和设计参数约束网络等。由它们形成装配设计模型。在概念和装配设计模型的基础上,详细建模完成产品的详细几何设计和性能设计。

产品设计的核心是资源重用。所使用的推理方法有:规则推理(RBR)、实例推理(CBR)和约束满足技术(CSP)。

#### 产品设计检验模块

产品设计检验模块包括需求检验、规范检验、合理性检验等。产品检验的第一要求是满足市场需求,具体表现在产品功能上,需求检验也是功能检验,这部分检验主要由设计师来进行,系统只是根据功能-结构关系来判断系统实现的功能。规范检验就比较具体,根据企业设计标准和规范,对设计师设计的产品进行检验,发现不符合规范的立即指出,并要求改正。如对于材料的选择,如果设计师任意选择材料,必然导致采购的困难和产品成本提高,所以必须在企业的合适的材料范围内选择,保证企业开发产品的低成本和高效率。合理性检验是对产品设计结果的检验,为保证快速的变型产品设计,必须符合一些合理化工程的哲理,如新零件数目的控制、结构的变化限制、产品可制造性等。

#### CAD 系统

CAD 系统完成产品设计的图形处理任务。用于变

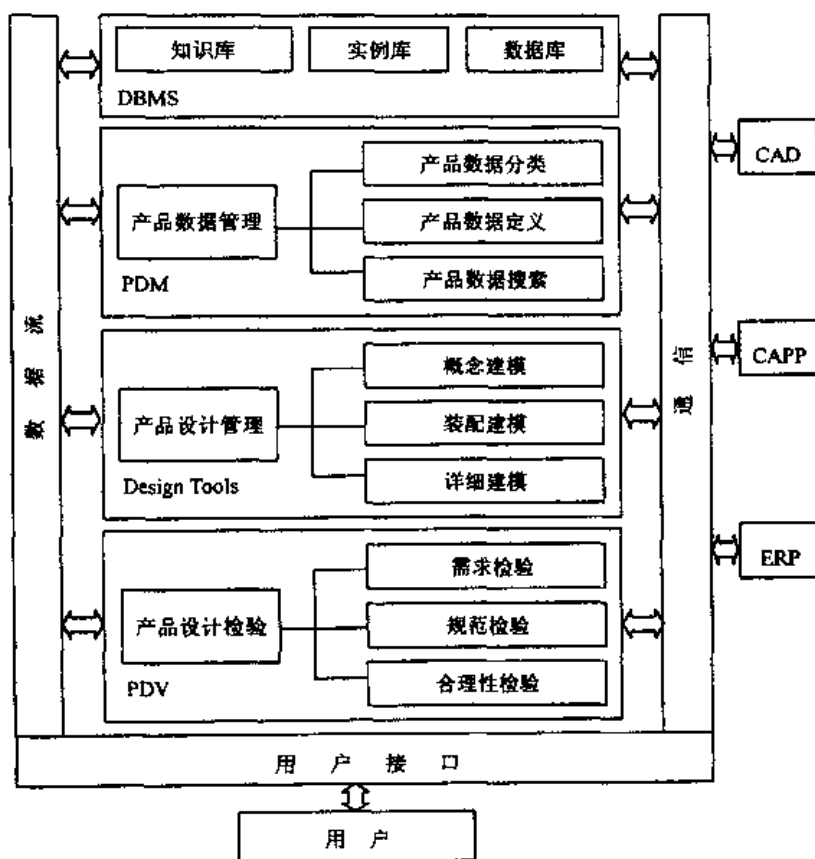


图 42.2-5 变型设计系统框架

型设计的 CAD 系统应具有一些与传统的 CAD 系统不同的特点。它以关系为核心,提供一系列工具来帮助设计师在设计对象之间建立和表达设计关系和设计约束,如图与图之间的关系、零件之间的运动联系、装配图的装配关系等。

知识库(KB)、实例库(CB)、数据库(DB)和数据管理系统(DBMS)

知识库存储设计规则、经验和设计知识(公理、定理、参数、图表等)。实例库存储设计参数约束、设计原型和实例对象。数据库中保存各种基础数据、标准件和产品设计过程的各种动态数据。数据库管理系统定义知识库、实例库和数据库的操作,是其他模块利用这些数据库的界面。

# 第3章 快速响应变型设计的关键技术

支持快速响应变型设计的关键技术包括:事物特性表管理、分类编码系统、产品资源管理和关系型CAD系统。

## 1 事物特性表管理

### 1.1 事物特性技术简介

在当今计算机化的信息社会,信息量日益增长,如何使纷繁的信息畅通无阻地流动,又如何及时得到所需要的有用信息,其中最关键的问题在于如何用一个统一的规范和格式来描述和记录这些信息。事物特性技术的运用,不仅可以描述对象的本质特性或者相关特性,还可以用于查询和管理对象及对象特性。

事物特性技术于20世纪70年代初产生于德国,德国利用事物特性技术建立国家级的信息流标准,并于1985年开始组织制订“事物特性表”系列标准。我国等效采用德国DIN标准,发布了国家标准GB/T10091.1《事物特性表定义和原理》及其系列标准GB10091.X。

事物特性技术适用于描述、限制和选择标准的和非标准的、物质的和非物质的、相似的事物对象或对象类。此外,事物特性表还支持文件中的数据归档、存贮和交换。在计算机化企业里,事物特性技术广泛应用于企业的各项活动,特别是在计算机辅助工程中,用来描述、存贮、查询和再现事物的特性信息。如在计算机辅助设计(CAD)领域,利用事物特性技术,描述原材料、零部件和产品的特性信息,并在设计开发过程中,查询、识别和比较有关材料和零部件的特性;在计算机辅助工艺编程(CAPP)中,利用对产品 & 零部件的特性描述,进行加工、装配和检验计划;在企业资源计划管理(ERP)中,对生产计划、能力计划、库存管理以及采购、销售和财务的各种特性信息进行描述和记录,以获得最佳生产、质量和成本控制;此外,也可用于其他系统的信息检索和处理,如图书、档案管理、商品市场信息、证券行情、电子数据交换(EDI)等。

### 1.2 事物特性

#### (1) 特性的概念

上述国家标准《事物特性表定义和原理》对特性(characteristic)的定义是:特性是从对象组中表征和区分一对象的决定性的性质。

任何事物的矛盾都有两个方面,即矛盾的特殊性质和一般性质,这种矛盾的特殊性质决定了一种事物区别于其他事物的特殊本质,这种事物的特殊性就是事物特性。对事物特性的认识,要经过从特殊到一般,从一般到特殊的反复循环,并要辩证地理解个性和共性的关系:事物的共性存在于一切过程的始终,共性存在于个性之中,无个性则无共性。那些描述和区分对象的信息,称之为特性。任何事物,无论是有形的,还是无形的,都可以用特性来表征和描述。

#### (2) 特性分类

特性分类的基本原则:

●科学性:按照不同的分类标准,可对特性进行不同的分类,特性分类的科学性主要在于分类标准的科学性。科学的分类标准,应能准确、合理地地区分特性。

●实用性:特性的分类应满足不同的应用需求,具有一定的针对性。在不同的应用场合,具有有效性和适应性。

●相关性:相关性包括特性相关和特性描述对象相关。

根据上述特性分类的基本原则,可对特性作如图42.3-1所示的分类。

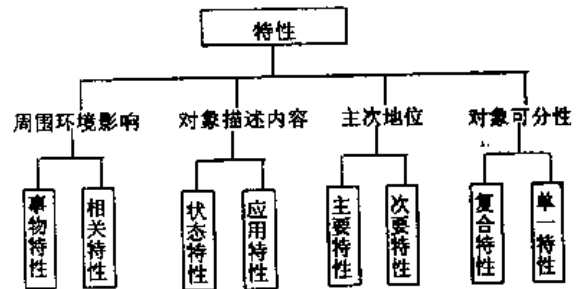


图 42.3-1 特性分类

全面而又准确地描述和识别一个对象(组),不仅要提取该对象(组)本身所固有的性质,还要对其生存空间中相关的特性进行描述。按是否受周围影响,可将特性分成事物特性(article characteristic)和关系特性(relation characteristic)。事物特性也称固有特性,是描述对象不受周围环境影响的特性。事物特性是决定一个对象的本质特性,是对象的固有性质,如人的性别、身高,苹果的大小、颜色,齿轮的齿数、模数等。如某一对象事物特性的特性值发生变化,则会导致该对象成为另一对象,而且任何一个事物特性的改变,都会引起对象

的变化,如改变齿轮的齿数,则形成另一齿轮,改变棒形半成品的截面,即形成另一半成品。关系特性是描述对象和周围环境的联系的特性。关系特性包括生产过程、起始值、生产费用、交货人、订购量、仓库地址、仓库存放费等。如苹果的成本、售价,调节弹簧的应用范围、装配高度、夹长、试验种类及文档标号等。关系特性的特性值的变化不导致所标识的对象的变化,比如苹果不管卖多少钱,还是这个苹果。某一对象的事物特性可能是另一对象的关系特性。如图 42.3-2 所示,对象 1 为齿轮,对象 2 为齿轮的市场情况,特性成本和售价在对象 1 中是关系特性,而在对象 2 中为事物特性。

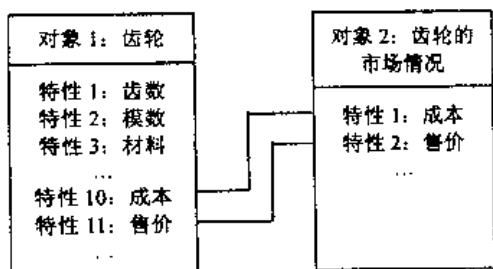


图 42.3-2 事物特性和相关特性的转换

按特性对对象描述的内容不同,可分为状态特性(characteristic of state)和应用特性(application characteristic)。状态特性是描述对象的状态的特性,如原材料、尺寸、表面状态、重量,苹果的大小、颜色,齿轮的齿数、模数,均属状态特性。应用特性是描述对象的功能和需要的特性,如连接尺寸、位置要求、能量要求、气候、耐热性、连接电流、容量。苹果的存储条件等均属应用特性。

按特性在对象中的地位,可分为主要特性(primary characteristic)和次要特性(secondary characteristic)。主要特性是对象描述中占主要地位的特性,一般说来主要特性基本决定了对象的性质。次要特性是在对象描述中占次要地位的特性,如齿轮的供应商和价格等。当检索一个对象时,一般先看主要特性,一旦主要特性不能满足需求,那么该对象就不是所要查找的对象。比如在采购齿轮时,如果齿轮的齿数和模数不符合要求,其价格再低,质量再好,也不能购买。特性的主次划分并没有明确的标准,它与对象的生存空间和应用场合有关。主要特性对次要特性具有支配作用,这种主从关系的数学描述如下:

$$P_i = f(S_i)$$

式中  $P_i$ ——主要特性的特性参数;

$S_i$ ——次要特性的特性参数;

$f()$ ——主要特性对次要特性的控制函数。

按特性是否具有可分性,可将特性分为复合特性(complex characteristic)和单一特性(single

characteristic)。复合特性是在共同概念下综合数个单一特性的特性,如绕组尺寸(绕组形式和直径)、电气设备和保护方法、应用等级、螺栓质量等级(尺寸准确度和表面粗糙度)、原材料(化合物、热处理、物理性能)。单一特性是不可分或不必要细分的特性。复合特性是由单一特性组成的,在对对象进行描述时,复合特性原则上应进一步分解成单一特性。如图 42.3-3,在对螺栓进行描述时,对螺栓的螺纹又作了一些描述,这样,对螺栓的描述更加完整。

代码	A	B	C	D	E
名称	螺栓	杆长	直径	头高	长
单位		mm	mm	mm	mm

A01	螺纹形式
A02	螺纹直径
A03	螺距
A04	

图 42.3-3 复合特性的分解

### (3) 事物特性的表征相位

为了从不同的要求、角度和层次对特性对象进行定量、定性的描述,应定义事物特性的表征相位。相位是指人类考察、试验、描述、表征事物特性的某个特定涵义的方面,从不同的相位去考察、试验、描述和表征事物所具有的个别结构、性质、功能等,可全面地认识、判断和识别事物。

●意相位:指事物的概念、定义,包括语言、文字、术语、代号和信号等,是对事物的总体表达。

●实体相位:指事物的客观形态、实体的一般性质及其长、宽、高、半径和面积等。

●质量相位:对事物本质进行表征的相位。

●功能相位:事物对客体产生的功能所呈现的特性。

●行为相位:对生产劳动与组织、社会经营行为中人与物、人与群体或群体之间的事务行为的描述。

其中,质量相位和功能相位是通过科学的方法取得的,称为方法相位。

### (4) 事物特性链

任何事物都不是单独存在的,事物间的相互依存和事物特性的相关性,形成了事物特性链。事物特性链是以不同形式的接口,将不同的事物特性连接起来描述事物间相结合关系的一种方法。按不同的接口形式,分为实物链、行为链和系统链。

●实物链:对实物等客观对象间结合关系的描述,其接口为客观实体的硬接口,如零件间的装配。

●行为链:对人与人、人与物的事务管理等抽象对象的行为关系的描述,其接口为管理行为等接口,如生产管理和计划中的各种关系链。

●系统链:系统链指系统间的结合关系。

特性接口是保证特性对象结合的技术条件。特性对象结合后,形成新的对象,具有新的事物特性。只有正确的特性接口,才能形成有效的事物特性链,组成新的组合对象。如波音 747 有 450 万个零部件,配套产品 2400 多项,技术标准和规范 2400 多项,如此庞大的系统工程,如果不保证事物特性的一致性和特性接口的正确性,其结果可想而知。

### 1.3 特性描述的对象

对象(object, article)是可用特性值描述的物质或非物质概念。对象可以是所接触到的每一实体,例如,汽车、人、计算机、零件、半成品、工具、书本等。凡是接触到的事物都可以称之为对象。对象也可以是数据结构和概念模型,例如,数据结构中的队列、堆栈等;概念模型中的实数、矩阵以及国家、车间,社会活动中的组织、服务等。对象还可以是处理过程,例如,工业系统的控制过程、原材料的使用过程、化学物质的变化过程、股票价格的变化过程、企业生产管理过程等都称为对象。

从上述不同的对象可以看出,对象是现实世界中一类具有某些共同特性的事物的抽象,如汽车对象是各种各样汽车共性的抽象。对象包含许多信息,这些信息可以用特性进行标识和描述,如队列对象含有队列长度、队列指针、队尾指针等特性,苹果对象具有大小、颜色和味道等特性。

从对象的定义和描述可以看出:特性所描述的对象可以是事,如各种管理事项,也可以是物,如各种物资;可以是具体的物质,如各种产品、人等,也可以是抽象的概念,如有关政策、金融、海关等信息。对同一对象的描述可以因不同的目的而有所不同,如对盒式录音带,可以对机械性能、电磁性能和音响特性等进行描述,也可以对所录信息,如节目名称、演奏或演唱者、指挥者、录制日期、发行单位等进行描述。

对象是现实世界中一类具有某些共同特性的事物的抽象。复杂的对象可以由相对简单的对象以某种方式构成。对象可有多种状态,但状态间必须是可逆的。如已加工的零件和新的未加工件因不可逆就认为是两个不同的对象。根据这些对象的共性和个性,可将对象分成不同的层次。

### 1.4 事物特性表(Layouts of Article Characteristics)

事物特性表是组合和排列对象组(group of articles)的事物特性和关系特性的表,所谓对象组是指一定的同族对象的组合。换句话说,事物特性表是一种把事物的特性描述出来,并统一规定存录和显现模式的信息标准。它提供了对象信息的描述、存放、查询检索、显现的标准格式,不同信息处理系统均可对其进行操作,使得异地、异构的信息系统间可以进行信息交换和传输。由此可见,事物特性表可用于概括地描述、限制

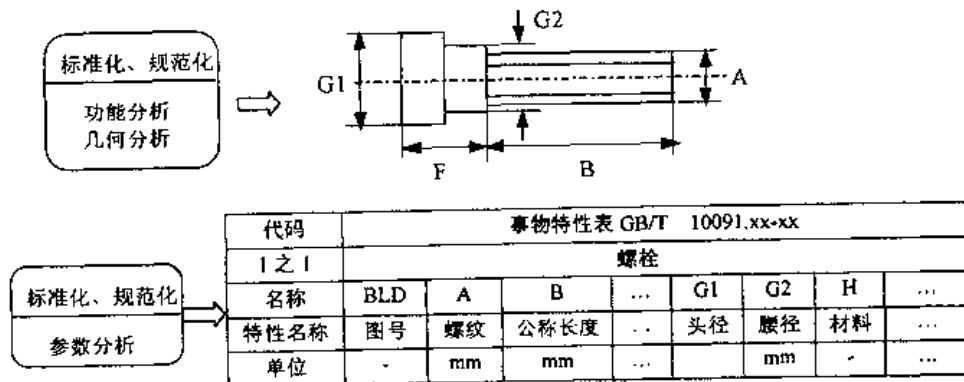


图 42.3-4 螺栓的事物特性表

和选择标准的、非标准的,物质的和非物质的以及相互近似的事物对象。将其用于 CAD 中,可规范化地支持 CAD 数据的归档、存储和交换。表的每一列表示相关的零部件的特性,表的每一行代表相关零部件类中的一个实例。图 42.3-4 是螺栓的事物特性表,GB/T10091-1 对事物特性表的组成和结构作了规定。

一般说来,事物特性表描述原则有以下几点:

(1) 覆盖面尽量广

一个事物特性表所描述的事物对象的种类应尽量

多,覆盖面应尽量广。也就是说,一个事物特性表应尽可能表达出对象类中有相同或相似特性的对象。如德国 DIN4000/23 部分《半成品》所表示的“产品外形特性”,其中的“板材外形特性”既表示了钢板、铝板等金属板材,又表示了木板、塑料板、橡胶板及其他板材。

(2) 目的明确

对对象的描述应根据应用场合和目的对所描述的特性作不同的选择。如对零部件,设计人员主要关心它

的技术性能;计划、供应和物资管理人员则主要关心它的货源情况、货源保证、合同和合同执行情况及价格等;而车间管理人员,除技术性能以外,还关心(对自制件)所用工装、模、卡、工、量具等,即一个事物特性表应根据不同的目的和要求,编成用多行或多分表表示出的具有综合内容的事物特性表。

### (3)力求简化

对一事物对象,应根据一定的需求只描述其必要的和充分的特性即可,不必描述出全部特性,即“简而充分”,所谓“简”就是给数据处理用户提供其必要的信息即可,避免对象描述过于繁琐和详细,所谓“充分”就是描述对象的特性应能分清两个不同的对象。

从事物特性表的应用范围,可看出事物特性表不仅描述的是事物特性,而且在有些应用场合将相关性描述出来也是必要的。如根据生产计划编制人员的需要,对产品的价格和货源进行描述。从变型设计的角度讲,事物特性分为零件特性和部件特性两类,特性种类主要分为:产品标准特性,主导特性,补充特性,功能特性,算法特性,分类特性,属性特性。事物特性表对企业产品资源的规范化、标准化具有重要的意义,DIN4000/4001对CAD设计中用到的所有标准件和企业通用件进行了描述,根据标准,标准件和企业通用件分成以下几组:事物特性标准(标准应用范围,支持企业标准);分标准(对象组);分表(组件族);分图(组件)。事物特性表管理包括创建、维护事物特性表,对象组、组件族、组件的检索和选取等功能,不仅管理简单件,并且用事物特性控制简单件连接,形成装配模型。

## 1.5 事物特性技术及其特点

事物特性技术是用事物特性和事物特性表的方法来描述和记录对象或对象类的技术。事物特性技术的运用,不仅可以描述对象的本质特性或者相关特性,还可以用于查询和管理对象及对象特性。归纳起来,事物特性技术具有以下几个特点:

### (1)抽象性

事物特性技术分析表达的并不是某一个具体的对象实体,而是对现实世界中一类实体或事物进行抽象的结果,也就是说事物特性技术是对对象或对象类的抽象描述。抽象既是一个过程,又是一个结果。客观世界纷繁复杂,抽象的过程可以从杂乱无章的事物中提取共同的信息,找到共同的运动规律;事物特性是对一类事物抽象过程的结果,当找到共同的信息和运动规律后,要对对象进行描述和定义,这个过程就可以用事物特性技术来完成。

### (2)继承性

继承性是面向对象技术的重要特性,与此相对应,事物特性技术也具有继承性。事物特性的描述对象可以按相似程度划分为不同的层次,相邻的对象类具有父子关系,子类继承了父类的属性。利用特性的继承性,可以帮助我们认识事物,并可用来区分具有相似特性的事物,也可以简化对象的描述。

### (3)可比性

事物特性技术的可比性包含两方面的内容,其一是利用事物特性技术可以使特性描述的对象具有可比性;另一方面,事物特性本身也具有可比性。利用事物特性技术,可以区分一组对象,比如在挑选苹果时,往往要考虑苹果的大小、成熟程度以及表面质量等因素,正是由于这些特性不同,才决定了苹果的优劣。在查找一个所需要的对象时,首先要是在脑海中有这样一个概念,该对象属于哪一类,应该具有哪些特点,然后将这些特点与现有的对象的特性进行比较,找出特性相同或相似的对象。这样一个特性比较的过程,也是对对象的认知过程。

对信息的描述和表达是信息技术发展的关键问题。由于人们对自然语言可能有不同的理解,传统的自然语言描述方法就难以明确、清晰、无二义地描述特性对象,难以建立信息模型,因而给计算机对结构化自然语言的表达带来困难。近年来人们致力于形式化的描述方法的研究,但对过于形式化的语言,直接用户又难以接受。面事物特性技术是一种介于两者之间的信息描述方法,较好地解决了自然语言和形式语言的不足,是信息资源描述的重要手段和方法,它已经在产品类属模型的建模上得到了成功的应用。

## 2 分类编码系统

### 2.1 零件分类编码系统的基本原理和结构

影响一个编码系统性能优劣的主要因素之一是它所使用的代码的结构,这不仅影响到编码系统本身的功能,也对能否易于在计算机上实现产生影响。

从数据模型理论的观点看,编码系统的结构也对应着一定的数据模型。当今编码系统常用的结构有树式、链式和混合式结构。树式结构对应于层次模型,链式编码结构对应于线性有序的数据模型。

1)树状编码结构:相邻两个码位间具有从属关系,每个码位(第一位除外)都进一步描述了前一码位所给出的信息,树状编码在系统码位数相同的条件下包含的信息量大,但结构较复杂,必须对分类对象的本质有十分深刻的了解,常用于不太能变化的固定信息的描述。

2) 链式编码结构: 其每个码位各不相干, 仅包含其自身范围内的信息, 它比较容易建立, 而且随着要求的变化也易改变, 一般用于描述信息变化较多的场合, 易于在计算机上处理, 但在相同码位数的情况下能包含的信息量较树式结构少。

3) 混合编码结构: 其实质是将某些码位形成的树状编码串接在链式编码的排列中, 兼有树状和链式结构的优点, 因其具有灵活的表达方式而得到广泛应用。

如今为人们所熟悉的两大零件分类编码系统是刚性编码系统和柔性编码系统。

传统的零件编码系统一般都是刚性编码系统, 如 OPITZ 系统和我国的 JCBM-1 系统等, 码位长度总是固定的, 无法根据需要取舍, 特别在采用高项覆盖低项的原则进行编码时, 不能充分描述零件的各种信息, 从而无法满足 GT-CAD/CAM 的要求。

柔性编码系统是当今才发展起来的, 根据数据模型的原理, 它突破了固定的框架结构, 用多层来相当详尽地描述零件的特征信息, 采用层层深入的方式, 以满足不同场合、不同单位的需要。所谓柔性, 一方面是指码位长度的不确定性, 可以根据需要随意增减; 另一方面是指信息在码位中的位置次序的随意性。

依据柔性编码系统的原则, 我国开发了一个 SE/BYJ-CCS 柔性编码系统, 它打破了传统的固定式结构, 由主码(固定)、副码(柔性)和辅助码(柔性)三层组成。

第一层主码(固定)主要用于描述零件的主体形状和主特征, 具有固定码位长度, 反映了零件的基本轮廓形状, 以及主要工艺信息, 能满足对零件分类要求, 可以据此确定零件族类别和典型工艺路线。

第二层副码(柔性)详细反映零件的除主特征以外的其他特征要素, 或是对零件的某些特征根据实际需要进行详细描述。柔性码中每两个字符代表零件一个特征信息, 表现形式为一字符串, 这就避免了传统编码系统中各种信息排列方式的限制和高项掩盖低项所造成的代码含义不清的缺点, 当不具备这方面的特征时, 可以略去。这层码为 CAPP 提供了较为详细的信息, 据此可以确定零件工序的操作内容。

第三层辅助码(柔性)是对零件几何特征或特征之间位置信息的补充描述。之所以突出这一项, 是便于 CAD 系统能输出 CAPP 系统在进行工艺设计时所必须的特征间的相互位置精度要求或特征的形状精度要求。如果 CAD 系统的输出已能满足上述要求, 此码可以不设。

## 2.2 适用于快速响应设计的分类编码系统

为了进行产品信息交流, 实现高层次上的技术管

理的合理化, 需要选择合适的分类标准。本文采用基于国标的面向事物特性的分类方法。要充分利用产品资源, 就必须提供一套明确的零部件检索机制。分类除了使企业内部在完成过程中信息交换合理化外, 其重要性还在于使设计人员能迅速和广泛地取得已经设计或作为库存件的现成或相似件的信息。在新型设计和变型设计时, 这种重复件的应用, 属于企业对设计人员最重要的设计要求之一。

自从 20 世纪 60 年代初, OPITZ 分类编码系统在德国诞生到现在, 国内外提出了多种分类编码系统, 但没有一种系统被普遍采用。OPITZ 分类方法的优点在于具有不同的抽象平面, 特别适用于相似寻找, 而缺点是信息损失量特别多和不灵活, 这是由于把精确的信息描述减少到某个值域的边界上。

事物特性(DIN4000)补充了这种分类系统的不足, 或者本身也是一种重要的辅助手段, 以便存贮和重新找出对象物, 例如标准件或公司专用的设计件、外协件、工具和生产设备等。同传统分类方法相反, 它适用于明确的零件描述, 因为它把握了精确的事物特性。但仅以事物特性为基础所组成的相似零件寻找系统, 由于它的命名上的问题和缺少顺序系统化而不能作为机械制造应用的基础。

可见, 分类系统和事物特性方法各有优缺点, 对于分类系统的精确描述来说, 它总难以避免信息损失, 因为缺少对象的精确值。而事物特性方法则相反, 虽然只允许一级分类, 它却能够不受限制地实现精确描述。将这两种方案结合具有很大的优势, 把分类系统的优点和事物特性方法的优点结合起来, 来实现对象范围的分级和分级后对象的精确描述。采用这种方案, 可建立一零部件检索系统, 其中标准件基于标准化命名分类, 自制零件基于形状特征分类, 部件基于功能分类。德国在这方面的研究起步较早, 成效显著、经验丰富, 并制定了相应的标准 DIN4000、DIN4001 及 DIN66304。我国也有等效采用的国家标准, 建立基于这些标准的编码系统, 为高层次的数据交换与共享提供了基础, 也是采用先进制造技术和制造哲理的基础。

有人提出了一种基于 DIN4000(GB10091)的面向产品生命全过程的合理化工程编码方法, 它是直接面向图形的, 便于设计人员利用 CAD 系统直接检索到零部件的白图和它所对应的数据表。设计人员对数据表中的数据进行复制、修改、存入等操作后即可得到新零部件的数据表, 该数据与白图合成后即可生成新零部件的图样。工艺人员得到新的零部件图样后, 可调出该零部件的标准工艺过程卡, 或相似零部件的工艺过程卡, 利用派生式 CAPP 的方法生成新零部件的工艺过程卡。该编码由识别号、事物特性码和视图编号三部



分组或(如图 42-3-5)。

#### (1) 识别码

采用零件分类编码系统加流水号作为识别码,虽然零件分类编码与事物特性码中的小部分信息有一些冗余,但这两个编码合在一起可以覆盖整个产品的制造过程。

#### (2) 事物特性码

事物特性码由标准编号、分标准编号、分表编号和分图编号组成。

#### (3) 视图编码

##### ●图形种类

G类组件:是用较多整件和必要的A和B构件组成的组件,仅指尺寸和(或)产品标准中的组件。

K类整件:用于描述一个完整的零件,一般由一个

或若干个A类或B类构件组成。

B类构件:专用于某一类图形的通用元素。

A类构件:最基本的通用几何元素。

##### ●显示等级

M—符号显示;S—标准显示;E—扩展显示。

##### ●视图种类

1—轴测图,2—主视图,3—俯视图,4—左视图,5—右视图,6—仰视图,7—后视图,一个视图可有多种不同的显示和形成该视图的变型(例如,剖面显示和隐边显示)。

##### ●组装方式

当一整件有不同的几何图形时,应清楚地将它们表示出来,例如,所属零件的形状可在组装时通过变形而改变,或同时变更尺寸。

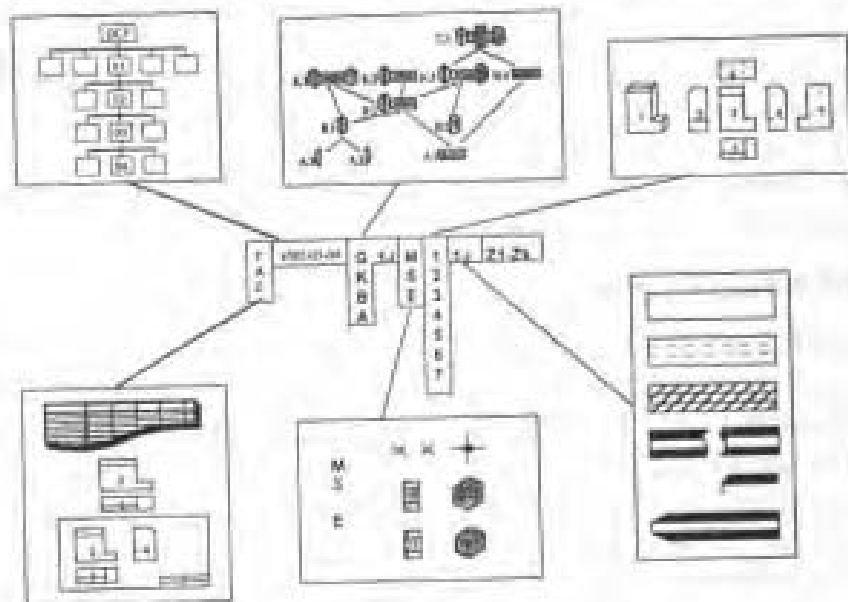


图 42-3-5 国内产品生命周期的编码方法

## 3 产品资源管理

### 3.1 产品资源管理的意义

企业成熟的零件、部件、产品都应看作是企业的宝贵信息资源。基于树状分类编码方法能够对产品资源的零件层、部件层、产品层进行统一的分类和管理,形成企业的产品资源库,建立产品、部件、零件之间以及它们与描述它们的文档、图样之间以及与供应商之间的关系,建立产品结构信息。关系型产品模型采用元对象模型统一、完整地定义和表达产品数据,使企业产品资源在其支持下,使用统一、规范和有效的定义。在这样的定义基础上,企业可统一管理其产品资源,使整个产品开发过程都可共享资源,快速的变型设计正是

在统一定义的产品资源的基础上,充分发挥信息技术的优势,将整个产品开发周期的所有信息集成在产品的设计结果,使后续的产品开发工作准备时间和周期大大缩短,整个产品开发周期也因此而缩短,同时产品成本下降。无疑这样的设计工作离不开信息环境的支持,因此,快速变型设计的信息化支持环境是必不可少的。另一方面,由于产品开发过程是一个流水线式的过程,在不同的信息平台上集成和交换各种数据,而且产品开发需要人员管理,包括设置设计人员的权限,设计数据版本管理,项目管理、图档管理等,所有这些客观上的需求表明,快速变型设计实施与合理、有效的设计支持环境有密切关系。

快速变型设计强调设计工作必须基于产品内在关系,强调利用这种关系来完整地表达产品信息,产品信

息是产品设计支持环境的主体,所有的产品设计支持环境的活动都是围绕产品信息展开的,如图档管理就是对产品视图信息的管理,版本管理是对产品历史信息的管理,产品数据的构成表明关系型产品模型是基于元对象模型来研究产品数据中的显式和隐式信息的定义和表达,这里的显式信息和隐式信息都是产品信息。产品信息和产品数据是有所不同的,产品数据是定义产品本身的内在和外在的信息,包括图形、装配关系、结构关系等,而产品信息是与产品设计结果和过程相关的所有信息。从这个角度考虑产品信息是包含了产品数据,同时还包括如设计手册、产品开发管理等信息。产品数据是统一用关系型产品模型统一表达,而产品信息是很难用统一的信息模型来定义和表达。

面向对象的关系型 CAD 系统,采用了变量化设计技术,具有表示和维持各种设计关系的能力,支持从概念设计到详细设计的新设计过程,也支持自底向上的变型设计过程。可以利用其强大的关系表达能力,配合事物特性技术来完成对企业已有产品资源的信息化重组。这些产品资源是企业最为宝贵的智力财富,关系表示和事物特性技术不仅是产品资源合理化重组的重要手段,也是保证产品资源能被充分利用的重要手段。

### 3.2 产品资源管理中的设计信息模型

设计过程中的产品信息可分为两类:设计前台信息和设计后台信息。设计前台信息(Design Foreground Information, DFI)一般指设计明确的结果,如产品视图、CAD 数据和模型数据等。设计后台信息(Design Background Information, DBI)是指在形成设计过程中所使用和产生的各种信息,这样的信息如设计意图、设计方法、设计标准等,往往随着产品结果的形成而消失,实际设计中这部分信息有助于理解和优化产品设计结果。本节重点是讨论产品信息的定义和管理,当然重点是在变型设计过程,围绕快速响应变型设计这一中心,对支持设计环境的产品信息进行讨论,以明确变型设计的基础和所集成的数据,所有这些都通过产品建模来实现。在动态产品模型支持下,便能实现自顶向下规划、自底向上实施的面向产品、面向零部件的变型设计。产品变型需要部件和零件的变型,部件和零件的变型决定着产品的变型。在产品总的装配规范的约束下,可以实现产品快速、合理的变型。

对于企业来讲,要使设计工程师充分重用产品资源,首先必须标准化、规范化企业资源,只有这样才有可能实现重用产品资源。标准化和规范化企业资源的第一步就是建立完备的资源命名体系,保证产品资源的唯一标识和快速检索。从上面的讨论中表明,作为可重用产品资源的命名体系必须满足以下条件:

- 产品资源命名必须具有唯一标识;
- 产品资源命名有利于反映其特性;
- 产品资源命名必须与传统习惯建立联系。

信息化的支持环境对其中的所有信息都进行计算机处理,另一方面设计工程师在产品设计过程中重用产品资源时必须要有不二义的定义和标识,才能在产品开发的全过程充分共享资源,才能在后续开发工作中快速实现信息转换,因此产品资源的命名必须具有唯一性,这与产品模型中元对象的名称唯一性是同一道理。另外产品资源命名要反映其特性,这样才能快速检索,否则对于变型设计实践是没有意义的。由于长期的历史原因,各企业都有自己的人工产品资源定义标准,用于设计资料的保管和使用,因此需要和原先的人工产品资源定义标准建立一定的联系。

在企业中一张图有图号,一份技术文件有其编号,这些就是人工定义的产品资源标识符号,这在于工操作的时代是符合其运作规律的,但在计算机管理时代就显得不符合规范了,如 135 柴油机的型号有长有短, G6135、G6135G1、G6135Z1、Q2A 都是 G6135 系列柴油机的型号,这样的命名一方面很难计算机处理,另一方面名字标识唯一性不强,容易产生混乱。在快速响应变型设计的支持环境中,识别码(Identify Number)、分类码(Classification Code)、辅码(Assistant Code)构成了产品资源的名字标识。识别码又称主码(Main Code),分为两部分,第一部分是资源分类码,这是简码,用于数据分段,第二部分是流水号,即产品资源在信息化的支持环境中的存储空间的标识,由计算机自动产生,便于计算机搜索、处理,主码不包含产品资源的具体信息内容,仅作为在信息环境中的名字。所有的产品资源,包括产品、部件、零件、标准件等在产品资源库中都视为统一的物料,按物料的性质分类存储。根据前节的内容,分类码由功能码、组码、族码、系列码组成,功能码是所属物料类的主要功能,功能码定义是根据预先规定的字典,在企业资源分类准备工作时,需将物料的主要功能逐一定义,如齿轮传动、连杆传动、飞轮平衡等,将这些功能定义好,编入数据字典。组码是实现该功能的一组物料,这是按抽象的功能划分的具有相似结构属性的一类对象,如齿轮传动中的圆柱齿轮传动可编为一组,这里的划分可以按国家、部委的有关标准,也可以按企业自己的标准进行,所有定义都必须预先编码。族码是在一组物料中按结构相同进行划分,如圆柱齿轮传动中可按展开式二级齿轮、分流式二级齿轮、同轴式二级齿轮、单级齿轮等结构分为若干个物料族。系列码主要是在同一结构下的每一个对象的编码,这样就可以唯一确定其在产品资源库中的位置。辅码则是说明产品资源库中的物料部分属性,协助计算机

进行有关的处理。辅码分为若干部分,分别说明如下:

产品资源的命名体系中还包括资源的名字,规划、整理和重用企业产品资源必须建立在一套规范的名字定义基础上,如果名字不统一,产品重用率是提不高的,所以在重组产品资源中,名字规范化具有重要意义。快速响应变型设计的支持环境中的名字由两部分构成,分别为名称(Name)、设计名称(Designation Name),所谓名称是经重组后的产品资源规范化的产品物料的名称,这些名称在数据字典中明确规定,它强调规范、统一,减少人为的多义性,设计名称是指在设计过程中传统的名称,往往与产品结构描述相关,这是指在产品的初次设计中定义的,作为一种传统的习惯的保留,不作为可重用的产品资源的名称。图 42-3-6 说明名称和设计名称的关系。

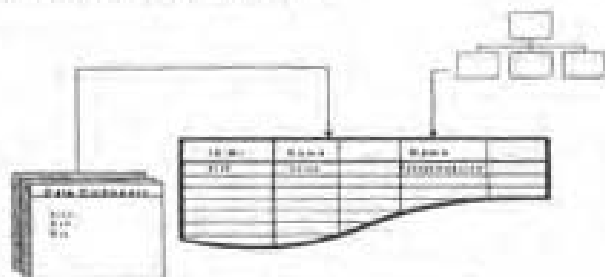


图 42-3-6 名称和设计名称的关系

产品在其历史发展的过程中,产品对象信息是动态、多变的,不仅在不同的产品开发过程会如此,在同一个开发过程中也会如此。对于如此变化的数据必须进行记录,有时常常要回溯原先的设计数据,所以产品的命名体系中应包括版本号。版本信息对于快速响应变型设计具有现实意义,版本信息包括两个方面:一方面是产品资源物料的升级信息,当该物料的新一代产生时,这种信息是传统意义上的版本信息,表示新一代替换对象已产生,在资源库中往往以生成新的物料标识号为标志;另一种是产品资源物料的修改信息,记录物料修改的历史,在资源库中并不产生新的物料标识号,有时也不生成新的视图,而是产生一个补充的信息,这实际上也是一种版本信息。

在关系型产品模型的基本属性中,对象类的显示属性是关于一个对象类的定义状态的形状在 2D/3D 平台所使用的视图的定义与表达。快速响应变型设计的结果最终要表现在可视化的模型上,而工程师使用产品资源也需要可视化的表达,尽管关系型产品模型的形状属性已描述了产品资源对象类的定义状态下的形状,但是合适的视图是物理地描述产品资源形状所必须的。这是由于有这几方面的理由,首先,产品资源的形状定义必须要用物理的描述表达清楚;其次,在更高一层的装配中需要物理的描述的支持;还有,定义状态的形状与重用状态的形状会有差异,如弹簧的自由

状态与工作状态,最后,在产品变型设计的后续过程中也需要取自定义状态或应用状态的形状的物理表达。所以,产品资源的一个重要标识就是可视化属性——视图标识。

视图标识中包括描述视图所代表的几何形状类别和描述视图自身属性完整的图形构件的代码体系。在本文的研究工作中,视图表达的是一个实例对象类的形状描述,自然它的分类属性在视图标识定义中是十分重要的。另一方面,前面谈及,不同应用场合需要的视图表达是不相同的,因此,产品资源的显示属性正是为了这样的特殊要求而定义的。视图标识包括以下几个部分内容:

- 视图标识的几何种类和标识号,它们描述了产品资源对象的基本图形单元及其标识;
- 视图标识的显示等级,它描述了产品资源对象在不同重用场合不同显示的简化定义与完整定义;
- 视图表达的视图代号的规定,这与重用中涉及的不同的视图方向直接相关;
- 视图表达的显示类别,产品资源对象的应用场合允许有一些非正常状态的显示(如局部剖、折断显示),另外当一个物料对象用于组装状态时会发生装配变型(如弹簧、铆钉),这些需要将视图显示定义为变型显示。

产品资源视图标识编码是依据合理化工程管理制定的,反映视图标识的内容,便于设计人员利用能支持快速响应变型设计的 CAD 系统,直接检索到零部件的白图和它所对应的数据表,包括所显示的各种状态,设计工程师对数据表中的数据进行复制、修改、存入等操作后即可得到新零部件的数据表,该数据与白图合成后即可生成新零部件的图样,这就是快速变型设计的特色之一。上节在适用于快速响应变型设计的分类编码系统中,就变型产品资源视图标识做了说明。

产品资源除了标识属性、视图属性,仍有许多其他属性,如标准级别属性、工作流属性、项目属性、权限属性等,这里主要讨论在关系型产品模型支持下的产品资源对象的属性表达和定义,这里的产品资源对象是指产品对象、部件对象、零件对象、特征单元对象,对于其他的产品资源对象由于与主题关系比较远,因而不作讨论,以下就用产品资源对象来代表上述的对象。

产品资源对象不是孤立地存在,而是与其他产品资源对象有机地结合在一起。产品资源对象往往本身就是由一些产品资源对象组成的,所以只表达产品资源对象本身是远不能满足产品快速设计的需求,还需要定义和表达产品资源对象间的关系。对象间的关系包括联接、聚合、归纳等类型。根据对象建模理论和关系型产品模型,定义如下:

联接是描述对象间相互结合的关系,这是一种对

等性的关系,即 is related to 关系,联接包括一对一、一对多、多对多等几种方式。

聚合是描述对象的构成、组合关系,是 is part of 关系,也是一对多的关系。它的反方向就是分解。

归纳是描述对象的一般性和特殊性的关系,是 is kind of 关系,也是一对多的关系,它的反方向就是继承。

用元对象和对象关系这些概念构成了面向对象的产品资源信息模型,也是关系型产品模型的组成部分。图 42.3-7 说明关系对象的表示方法,图 a 说明联接关系的四种方式,图 b 说明聚合关系,图 c 说明归纳关系,用这些方法可详细地定义和表达整个产品资源对象的所有信息。

尽管从理论上讲,可用元对象和对象关系完整地描述产品资源对象的信息,但由于产品资源对象是非常复杂的,其信息量是巨大的,所以如只用一层要描述清楚产品资源的关系是相当困难的,必须将产品资源分层。产品资源分成产品、部件(包括若干零部件组成的特殊组件)、零件、特征单元四层,除了产品层,其他层次又可以分成若干个小层次,具体分法视产品资源的复杂性而定。这样在每一层上的产品资源数量和关系有限,就能清楚地表达产品资源的信息。

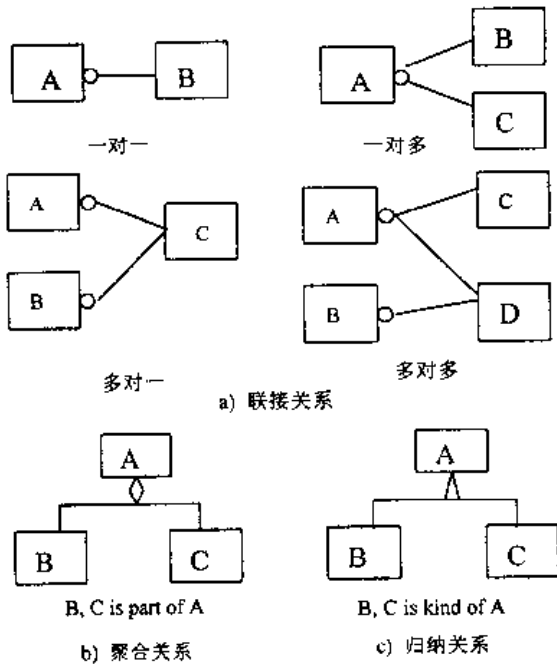


图 42.3-7 产品资源对象关系

### 3.3 产品信息管理

信息模型通常采用面向对象的信息表达方法,产品资源对象是信息模型的基本单元,从理论上讲,产品信息物理层的组织和管理也应采用面向对象方式,这就是目前面向对象的数据库技术方兴未艾的原因之

一。但从实践的角度出发,关系型数据模型由于其成熟的关系代数的支持和描述数据对象的直观性,使得关系型数据库成为商品化的数据库系统的主导,到目前为止,情况仍未改变。由于关系型数据库系统长期实践而形成的相当可靠性,使得在数据库领域中短期内仍无法出现成熟的面向对象的数据库系统,另外企业长期积累的产品信息基本上也都储存在关系型数据库中,如何转换也是一个严重的问题。所以本文从实践出发,依托关系型数据库系统,通过面向对象的数据管理机制,实现产品信息数据管理。

产品信息数据物理存储在关系型的数据库中,充分利用关系型数据库系统的数据关联性,在关系型产品模型的支持下,建立面向对象的数据管理机制。快速响应变型设计系统处理的数据,经过数据管理接口,按表的形式储存。当设计工程师需要这些数据时,通过数据管理接口将不同表的数据结合在一起,形成对象模型数据结构,返回到快速响应变型设计系统。图 42.3-8 是产品信息数据管理模型,底层是物理层,包含操作系统和数据库,其上是数据库管理系统,如 Sybase、Oracle、MS-SQL Server 等和产品数据模型;第三层是面向对象的数据管理机制;第四层是最高层,即应用层,直接面向应用软件。

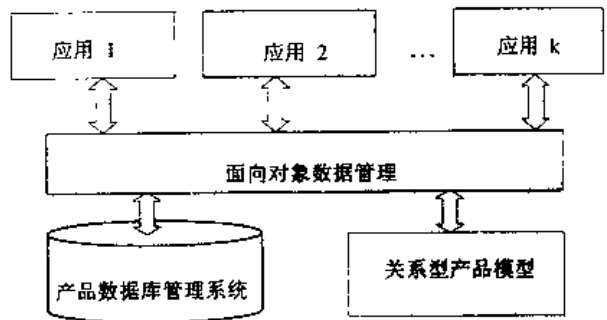


图 42.3-8 产品信息数据管理模型

本节首先研究产品资源对象的信息如何在关系型数据库中存储,数据管理和检索速度与库结构密切相关,因此库结构设计的好坏是产品信息储存有效性的前提。在前节中已说明了产品信息的类型、内容,这些信息以产品资源对象的属性表示,这些属性的内容是不固定的,但关系型数据库中的基本单元是表,表的结构是固定的。本文中所有产品资源对象的属性表都按抽象层次结构中的族来划分,同一族的产品资源具有相同的结构,所以也具有相同的属性参数种类,这样就可以固定表结构。每一个产品资源对象都对应一条主属性记录,其数据项包括标识码、分类码、几何属性、参数属性、视图属性等。表 42.3-1 是某零件的主属性记录的数据项结构。

表 42.3-1 零件主属性表

数据项名		数据项内容
标识码		产品资源标识码
几何属性	长度	
	定位尺寸	
	直径	
	角度	
参数属性	表面精度	
	尺寸公差	
	位置公差	
视图属性		产品资源视图编码
分类属性		产品资源分类编码
...		...

表 42.3-1 说明了对象属性值基本上表现在产品资源主属性表中,其他还有一些辅助表,如产品资源分类表、产品资源视图表、产品资源结构表、算法表等。这些表补充说明产品资源的信息,如属性之间的计算公式、产品资源的图形、产品资源的功能结构对应关系等。关系型数据库系统通过主属性表将所有这些辅助表连接在一起,构成完整的产品资源信息表。图 42.3-9 表示产品资源在关系型数据库中的存储结构。

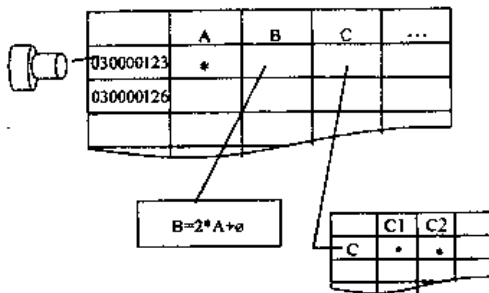


图 42.3-9 产品信息储存结构

在产品资源的存储结构中,需要强调的是产品资源的分类。产品资源对象有两种分类方式:结构分类和抽象分类。这些分类特点如何在关系型数据库系统中体现,这在上面的说明中已经涉及,也就是所有产品资源对象的主属性表是按族建表的,一个族就对应一张表,关于族对象的抽象描述就在该表的第一项。根据这两种分类方法,在产品资源数据库中构筑了两张索引表,分别表示这两类分类方法。结构索引表的数据项分别为父项和子项,由于产品资源的结构是多对多关系,所以在结构索引表中的数据是有冗余的,但这样表达产品资源结构是最清楚的,可以上下两个方向搜索,用 SQL 语句就能快速查询到所有的结构。结构索引表中的父子项内的数据是产品资源对象的标识号。抽象索引表的结构与结构索引表结构相似,但产品资源对象抽象层次结构是一对多关系,表中的数据项是上下级抽象对象标识号。最底层的就是上面讨论的代表产品最后结果的产品资源对象标识号。对于族对象来说,在产品资源库中每一主属性表代表一个族对象,其抽象

标识号也就是表号。在功能结构表中描述了功能对象和组对象,每一功能对象和组对象都有其固有的对象标识号,也是抽象标识号。有了两个分类索引表,设计工程师就可方便地根据两种路线搜索所需要的产品资源对象,产品结构索引表实际就是产品结构明细表,产品设计工程师能根据某一产品资源查询到它所构成的所有子产品资源对象,也能查询到所有调用过该产品资源对象的产品资源对象。这对后续的工艺、生产安排等开发活动提供了有力的查询工具,同时对设计工程师也是有力的工具。产品抽象索引表是作者提出的适应快速响应变型设计的产品数据索引,基于这样的索引,产品资源对象在变型设计的全过程都被记录下来,便于设计工程师理解设计的结果,也便于以后的变型设计的规范化。

在快速响应变型设计系统中,关键问题是快速检索到这些产品资源对象。每一个产品资源对象都有许多的属性定义和方法定义,每一个属性可以由一个数据项表达,也可由一个数据文件来表达。对于如此繁杂的产品资源数据,如没有合适的检索策略,难以有效地检索需要的资源信息,用于支持快速响应变型设计的支持环境也失去了应有的作用。在制造企业中,不断地无序产生新的零件和产品资源的命名混乱,相当程度是因为检索所需数据的速度和效率都太低,设计工程师不愿意搜索企业的产品资源库,宁愿自己重新设计。产品资源标识号是具有唯一性的标识符号,用其检索的速度是很快的,但产品资源标识号是计算机内部码,产品设计工程师很难也没有必要了解和记住这些流水码,因此整个产品资源的检索策略必须建立在可行的基础上。产品资源的检索策略包含两个方面的含义,一方面是根据初始的输入参数形成搜索关键词,另一方面分析搜索的结果集,找出符合问题的最适合的结果。

### 3.4 产品后台信息的定义与管理

本节前已说过,产品信息包括两个部分,一部分称为前台信息(DFI),而另一部分称为后台信息(DBI)。从概念的定义出发,DFI 主要是指产品最终结果,表达“What”的含义,而 DBI 主要指随设计过程而生成的信息,表达“how”和“why”的含义,如设计需求、设计说明、设计意图、设计策略、设计规范、设计标准、历史版本、过程记录、工程分析信息和仿真信息等。实际上 DFI 和 DBI 之间的分界线是很难划分的,如设计意图常常表现在设计的最终结果,另外产品资源对象间的关系也常常体现在产品最终结果,所以本节讨论的 DBI 所包含的信息有可能与前述讨论 DFI 的信息相重叠。

关系型产品模型强调产品信息之间的关系,认为

产品资源重组和快速响应变型设计都以产品信息间的关系为基础,只有通过关系才能将各种信息融合为一个有机的整体。对企业实践的考察发现,由于企业内部各种种类繁多、相互关联的技术文档没有采取很好的措施进行定义和管理,造成技术文档编制周期长、更新速度慢,还有技术文档内容不统一、不规范、不完备、不准确,会出现几个部门的文档都不一样,而且都是有效文档,可以想象到在如此情况下是不能保证企业快速响应市场的,甚至不能保证产品质量。而 DBI 是设计过程中生成的有关产品功能、制造、设计等各方面的信息,是与产品最终结果密切相关的信息,它的重要性体现在以下几个方面:

首先, DBI 保留了设计工程师的设计经验,在快速响应变型设计过程中,设计工程师可以从原先的产品资源的 DBI 了解到以前的产品设计历程,从而掌握该类产品的的设计经验。这对培养新的设计工程师尤为重要。

其次, DBI 说明产品设计环境和意图,使后来的设计工程师可以迅速了解产品设计情况和设计者的意图,这对产品优化和改进有相当的好处。快速响应变型设计重在产品改进设计上,可有时必须全面了解产品设计思路后才可进行变动设计,否则容易出现错误,尤其对于那些复杂的产品系统。如修改产品的零部件时,必须了解哪些部分是相对独立的,若进行修改不会影响其他零部件,哪些则是与其他零部件关联的,牵一发而动全身。在企业考察中这种情况有时需要主任设计师来决定。

最后, DBI 提高了实现并行工程的基础。有了 DBI 的帮助,即使当设计工程师尚未完全完成设计时,他(她)仍可将设计目标、设计意图、变型原则等相关信息和设计半成品向后续产品开发部门传递。制造工程师或工艺工程师则根据 DBI 就可以比较完整和准确地

了解所设计的产品结果,能先于设计工作结束前开展产品开发的后续活动,如工艺制定、制造安排、材料准备等。DBI 可以提高产品开发合作的有效性。

尽管 DBI 有如此的重要性和对快速响应变型设计的支持,但从上面的分析中了解到, DBI 包含的信息种类繁多、内容格式不一致,很难模型化和格式化。当然也有部分 DBI 是有规律和格式的,如日常设计工作中经常使用的设计规则手册、计算手册、实验报告等。因此如何表达 DBI 以及如何将 DBI 与 DFI 连接是 DBI 研究的重点。

在讨论 DBI 表达之前,首先要认识产品数据管理(PDM)所涉及的数据和 DBI 之间的关系。目前 PDM 仍未有明确的定义,一般认为 PDM 是以软件为基础的技术。它将与产品有关的信息和与产品有关的过程集成到一起,有效地将产品从概念设计、计算分析、详细设计、工艺流程设计、加工制造、销售维护直至消亡的整个生命周期内及其各阶段的相关数据,按照一定的数学模式加以定义、组织和管理,使产品数据在其整个生命周期内保持一致、最新、共享及安全。与 DBI 包含的信息相似, PDM 所处理的数据范围也很广,包括产品图档、制造文件、工艺文档、工作小组、 workflow 规则、结构明细表等。同时 PDM 更注重这些数据的过程控制,如设计的修改需要得到授权,并对其他使用者和相关者发出通知。当修改完毕,收回设计师的修改权,并将修改结果经过评审以得到确认后重新刷新数据库记录。从上面的分析说明,当前 PDM 系统主要处理 DFI 及其相关信息,这些相关信息部分与 DBI 重叠, DBI 中许多信息是当前 PDM 系统不处理的,如设计意图、设计条件、设计手册。当然 PDM 是一个可扩展的系统,将来或许会包含这些信息。图 42.3-10 表示了 DBI 和 PDM 处理数据的关系。

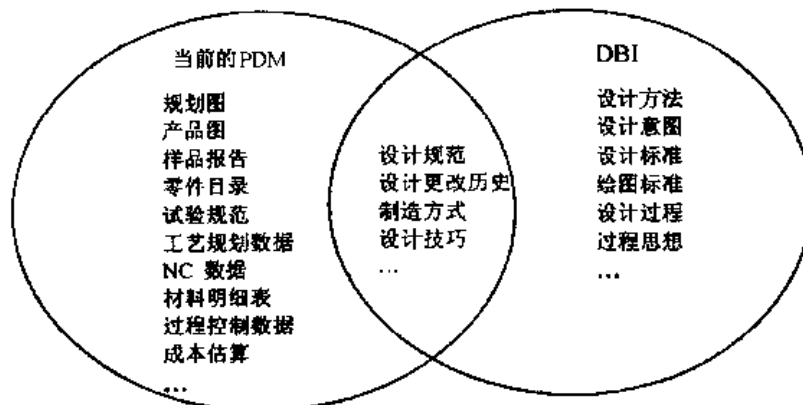


图 42.3-10 PDM 数据与 DBI 的关系

DBI 包含了许多知识的表示,所以 DBI 采用框架式表示法,每一条 DBI 占有一个槽。由于 DBI 同时包含动态变化的信息和相对静态的信息,因此 DBI 修改

频率比较高,而且可能会扩充。采用框架式方法表示比较灵活,增加 DBI 可以增加槽,修改也很方便。由于 DBI 包含了许多阐述性语言,如“法兰底部孔为不通

孔”、“箱体盖不能与其他零部件接触”等,为了避免不同的工程师采用不同的语言描述,而造成 DBI 的多义性,可模仿计算机编程语言定义的工程描述语言,用于描述工程情况,如设计规则可用“If ... then ... end If”的形式来表示,对于明细表中父子关系可用“A ⊂ B, A, BC X”的形式表示。这里 A 是 B 的子项, X 表示产品资源集。

DBI 另一个重点问题是如何与 DFI 相关联,否则 DBI 也没有意义。这个问题可分为两方面问题,一方面是 DBI 如何产生,另一方面是 DBI 如何应用。

根据 DBI 的定义,显然 DBI 是在产品设计过程中产生的。DBI 产生过程首先是信息收集过程, DBI 中的部分信息是在设计活动前预先定义好的,如产品设计标准、绘图标准、设计规则等。这些信息一般都以手册等文档形式出现在设计工程师面前,这种形式存在查找不便、关联不强等缺点,而且由于 CAD 系统普遍使用,这些信息再以手册等文档形式出现就显得不合适了,限制了设计速度和效率。这部分信息可以一定的格式输入到数据库中,形成通用的 DBI 数据库。对于大量的与具体设计相关的 DBI,产生的方式就有所不同了。在关系型产品模型的支持下,关系型 CAD 系统在设计的过程中会记录许多设计对象间的关系。在快速响应变型设计过程中,设计工程师检索到原型,同时也把原型所属的 DBI 调入,再根据 DBI 和设计需求进行变型设计。对于改动部分则记录改动过程和关系,同时设计工程师还输入部分信息,如改动意图、适用环境等,这样旧的 DBI 加上新的信息就作为新的变型产品的 DBI。接着对所收集的信息进行一定程度的模型化,按框架表示方法对所收集的信息处理,提取关键词,按预先定义好的格式填写入每一槽中,最后存储到数据库中。在 DBI 表示模型中每一个框架有一指针,指向 DFI 中的产品资源对象,这样通过指针建立了 DBI 和 DFI 之间的联系。这个联系是基于关系型产品模型, DFI 中产品资源之间的关系数据存放在 DBI 中,当快速响应变型设计系统在重定义或重用 DFI 中的产品资源时,必然与 DBI 发生联系。

## 4 关系型 CAD 系统

支持变型设计的 CAD 系统要支持包括概念设计在内的全过程,要存储设计者的设计知识,表达设计意图,并支持设计结果及设计过程的重用。

### 4.1 传统 CAD 系统的局限性

传统的 CAD 系统起源于计算机图形学,其智能定位于图样绘制,没有从设计本身的需求来考虑,大多停留在电子图板的水平。设计者用 CAD 系统来记录设计结果,设计

活动只发生在设计者的脑海中,当他们应用 CAD 系统的时候,设计差不多已经结束。主要表现在:

- 只是一个绘图工具而并非设计工具,不能帮助设计者定义设计关系和设计约束,更不能储存和保持设计关系。

- 不能充分利用已有的零部件,标准化设计思想无法在设计过程中得到贯彻和实施。

- 没有可变型的产品模型。

- 不支持设计全过程,只能完成绘图等对提高企业竞争力不很重要的工作。

- 缺乏智能性,只记录几何数据,缺乏工程语义信息,不能有效表达设计意图。

- 对产品缺乏完善的分类系统和检索机制。

几何造型是一个 CAD 系统所必需的功能,人们在设计时,需要将高层的功能需求转换为低层的几何数据,在转换的过程中,大量的语义信息丢失了。目前, CAD 技术已被广泛应用于机械行业,但传统 CAD 系统只注重表达几何信息,不能满足设计中知识表达和关系表示等要求。仅仅使用计算机来辅助完成图样绘制、零部件造型等技术设计和详细设计工作,已不能满足设计需要。方便设计者表达设计知识的新一代 CAD 系统成为发展的必然。

设计者应用 CAD 的基本过程是,对基本的技术对象进行构思,产生设计约束,然后把把这些设计构思、设计约束进行抽象,把高层的工程语义信息转变为计算机能够理解的低层的几何信息,建立计算机内部模型,然后在内部模型的基础上进行各种分析活动,如公差分析、运动模拟、装配分析等。

图形的绘制、编辑、显示等功能对一个 CAD 系统来说是必需的,但传统 CAD 系统的数据结构中仅保存图素的几何信息,而对图素间的各种约束关系没有记录。例如,一条线段的创建方式有多种:

- 两点间连线;

- 通过某点的水平直线;

- 通过某点具有一定长度且与水平方向成某一角度的直线;

- 过某点与另一直线平行的直线;

- 从一点到圆作切线;

- 两圆公切线等。

但无论何种方式,在系统的数据结构中保存的仅仅是线段的两个端点的坐标值。这样,有许多工程语义信息丢失了,在系统所记录的设计信息和设计者的设计意图之间存在着很大的差距。可见,要表达设计意图,有必要把对象的创建方法存储在系统的数据结构中,设计关系作为行为的结果被创建,例如相切、垂直等关系,在设计的过程中,这些关系能够方便地检

索、修改、更新和删除。在关系型模型中，系统数据结构记录的是对象的创建方式和创建依据，这样，图素类型、创建方法及构造参数都得以保存。系统按面向对象的方式组织相关的动态算法，当改变一个图素时，动态算法按数据结构中保存的关系，自动修改相关的其他对象，使其符合设计者的意图。可见，新模型是过程式的，而非纯结果式的。

如图 42.3-11 所示的图形，其生成过程如图 42.3-11b 所示，在传统 CAD 系统的数据结构中保存的是  $p_1$ 、 $p_2$  点的位置，圆  $c$  的圆心位置及半径，线  $l$  的起点及终点的位置。当其中任一对象发生变化时，其他对象保持不变，不能表达设计者对对象关系的定义。而关系型 CAD 系统数据结构中保存的却是各对象之间的关系， $p_2$  基于  $p_1$  生成， $c$  以  $p_2$  为圆心， $l$  过  $p_1$  且与  $c$  相切，当其中任一对象发生变化时，对象间的关系保持不变。过一点与圆相切这种方法使系统记录的是点  $p_1$  和圆  $c$  这两个参数，而非直线终点的坐标。

如果  $c$  的大小或位置发生改变，相切的关系不会受到影响，直线  $l$  将随着圆的改变而发生变化，但仍保持过点  $p_1$  与圆  $c$  相切的关系不变。这意味着改变的仅

是  $l$  的参数，而  $l$  仍保持为一切线。例如，当圆  $c$  的半径发生改变时， $c$  的尺寸发生改变，位置不变， $p_1$ 、 $p_2$  都保持不变，只有切线需要重新求解(图 42.3-12a)；当  $p_2$  发生变化， $c$  的位置发生改变，尺寸保持不变， $p_1$  保持不变， $l$  需要在此基础上重新求解(图 42.3-12b)；当  $p_1$  发生改变而其他保持不变时，因为只有  $p_1$  点是直接根据坐标原点创建的，而其他各个对象都依赖  $p_1$  点，因此它们只是随着  $p_1$  点偏移(图 42.3-12c)；如果  $p_2$  点不是相对于  $p_1$  点创建，而也是直接相对于坐标原点创建，则改变  $p_1$  点不会引起其他对象的改变，只是在此基础上对  $l$  进行重新求解(图 42.3-12d)。

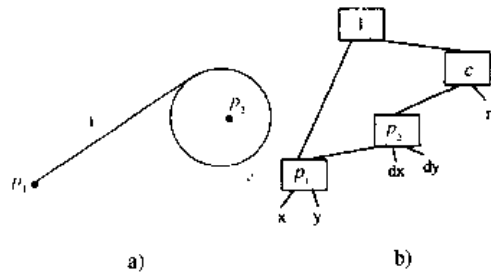


图 42.3-11 图素间关系及创建过程

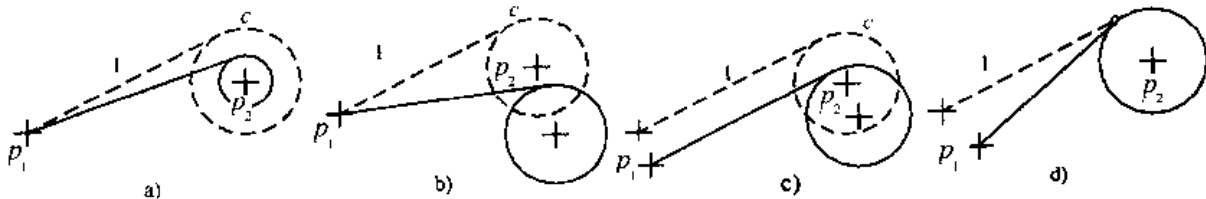


图 42.3-12 过一点与圆相切关系举例

### 4.2 参数化设计和变量化设计

现在，CAD 基础技术之争主要集中在以 PTC 为代表的参数化技术和以 SDRC 为代表的变量化技术之间的竞争，这两种基于约束的造型方式推动了 CAD 技术的革命。

参数化技术用顺序方法对约束方程进行求解，方程一个接一个求解，直至所有的变量被确定，有多少未知变量就要求有多少约束方程，并且变量之间不应耦合；变量化技术用平行方法对约束方程进行求解，方程组一般是非线性的、耦合的、非对称的，方程数目可以比变量数目少( $m < n$ )，也就是所说的欠约束情况，这时的解决方案就是有效解中的一个。

变量化技术是一种设计方法，它将几何图形约束和数学方程耦合在一起联立求解，以图形学理论和强大的计算机数值解析技术为设计者提供约束驱动能力。参数化技术是一种建模技术，应用于非耦合的几何图形和简易方程式的顺序求解，用特殊情况找寻原理和解释技术，为设计者提供尺寸驱动能力。我们可以把两种策略看作对一组约束方程求解，方程组由变量(未

知量)和约束(问题的边界条件)组成。一般，方程表达式中有几何变量(如表达长度、高度的尺寸等)，几何约束(如平行、相切等)，工程变量(如流速、传动比等)。

两种技术最根本的区别在于是否要全约束以及以什么形式来施加约束。参数化造型的特点是：全数据相关，全尺寸约束，尺寸驱动设计修改。它彻底克服了自由建模的无约束状态，几何形状均以尺寸的形式而牢牢地控制住。变量化技术将参数化技术中所定义的尺寸参数进一步区分为形状约束和尺寸约束，而不是像参数化技术那样只用尺寸来约束全部几何。参数化技术的工程约束不直接参与约束管理，而是另由单独的处理外置处理；在变量化技术中，工程关系可以作为约束直接与几何方程耦合，最后再通过约束解算器统一解算。

由于参数化技术使用预先定义的尺寸模式和预先定义的求解顺序，求解稳定、可控，在商业化的 CAD 系统中很普遍；由于缺少求解策略使变量几何问题很难求解，因为，求解器必须解决下列问题：获得求解顺序，处理大量的非线性方程组，处理多解问题并且判断设计意图。



从数学方法上来看,变量化技术是参数化技术的一个超集,更一般化的变量化技术为设计师提供如下好处:

- 能处理部分约束问题,因而更自然、更有效率。

- 能处理更广泛的工程问题,包括耦合的几何和工程关系。

- 因为对设计约束的显式处理,能提供一个更可理解的方式来定义和修改设计问题。

由于参数化技术苛求全约束,每一方程式必须是显函数,而所使用的变量必须在前面的方程式内已经定义过并赋值于某尺寸参数,其几何方程求解只能是顺序求解;变量化技术为适应各种约束条件,采用联立求解的数学手段,方程求解顺序无所谓。参数化技术解决的是全约束下的几何图形问题,表现形式是尺寸驱动几何形状修改;变量化技术解决的是任意约束情况下的产品设计问题,不仅可以做到尺寸驱动,亦可实现约束驱动,即由工程关系来驱动几何形状的改变。

变量化设计的过程是一个类似工程师在脑海里思考设计方案的过程,满足设计要求的几何形状是第一位的,尺寸细节是后来才逐步精确完善的,设计过程相对自由宽松,设计者有更多的时间和精力去考虑设计方案,而无需过多关心软件的内在机制和设计规划的限制,这符合工程师的创造性思维规律。变量化设计系统的应用领域更广阔,除了系列化零件设计,变量化设计系统在作概念设计时特别得心应手,比较适应于新产品开发、老产品改型这类新颖设计。I-DEAS 软件采用变量化技术作为系统的底层,它的成功意味着所有成功的机械设计自动化软件将顺应这一历史潮流。

#### 4.3 适用于变型设计的关系型 CAD 系统

适用于变型设计的关系型 CAD 系统应做到:

- 便于设计者表达设计知识,并自动对这些设计知识进行记录,便于查询和修改。

- 使标准化、模块化、系列化的设计原则得以贯彻实施,使重用产品资源的思想得以实现。

- 面向设计全过程,保持概念设计、装配设计、零件设计的逻辑一致性,支持自顶向下的初始设计和自底向上的变型设计过程。

- 强调产品数据管理,对已有产品资源建立一套广泛的、实用的查找方法和系统。

- 将方法学跟先进的 CAD 技术(特征基设计、约束基设计)结合起来。

##### (1) 关系浅析

在设计过程中,设计者要表达各种各样的工程语义关系,如装配层次关系、配合关系、位置关系、联接关系、运动关系等。在 CAD 系统中,这些关系可以借助

抽象的几何及非几何对象间的关系来表达,要准确表达设计知识和设计意图,就要保证这些关系能够以一种完全、一致、统一的策略进行表示和维护。关系是整个设计对象内涵保持一致的基础,也是各种设计知识和设计约束的载体。通过关系可以建立包括设计师的设计思想在内的设计模型,从而使模型具有智能性。正确、合理地运用这些关系,可以建立机械设计中的概念模型、装配模型和零件模型。

关系是指几何信息间、非几何信息间以及几何信息和非几何信息间相互作用、相互影响的状态,关系包括组成关系、拓扑关系、数值关系三类:

- 组成关系:对于几何信息表达几何对象间的成员关系。如圆有圆心、半径两个成员对象,直线有起点、终点两个成员对象。

- 拓扑关系:例如两维空间中几何对象间的位置关系,如平行、垂直、相切、相交等。

- 数值关系:长度、角度以及变量间的数值关系,如圆的直径等于一条直线的长度。

在产品设计的同时,这些关系被记录和维持,即无需设计者多余的输入,系统自动捕捉一些重要的设计意图信息,在设计编辑或修改期间,模型的任一部分都可以自动得到更新。在关系型 CAD 系统中,对象具有智能性——每一对象都知道其自身是如何创建的,依赖于哪些对象,并且知道在得到更改通知后如何更新自己。关系模型也具有智能性,它知道它的对象是如何连接的,因此它能够以确定的顺序给确定的对象发出更新命令,以使它们保持正确的连接。

参数化设计系统和变量化设计系统,一般也被称为基于约束的设计系统,约束也可以分为工程约束和几何约束两类。在这些系统中,几何约束分为两类:拓扑约束和尺寸约束。拓扑约束有相交、平行、垂直等,尺寸包括长度和角度两种类型。可以看出,关系和约束具有某些相似之处,但它们却是两个不同的概念。关系表达的是一种面向过程的状态,约束是一个面向结果的静态概念,是维持对象间已经声明的关系的边界条件。它指某种关联下两个或多个信息间值的约定,是对可能的解的范围的一种限制;关系的工程语义表达能力强,可以表达各种对象间的关系,约束的抽象层次比关系要高,工程语义表达能力相对较弱;基于约束的 CAD 系统要对方程组联立求解,而在关系型 CAD 系统中,方程组被分解,并封装于几何或非几何对象之内,有效地减小了方程组的求解规模。在基于约束的 CAD 系统中,几何对象的地位是平等的,即系统的数据结构开始为一无向图,然后基于自由度分析,再转化为有向图,而在关系型 CAD 系统中,对象一开始就以有向图的形式保存在系统中;在基于约束的 CAD 系

统中,一般需要先标注尺寸,然后驱动图形的修改,对规模较大的图形,可控性较差,在关系型 CAD 系统中,只需改变想修改的对象的参数即可。关于关系和约束的定义见下面的叙述,实际上约束和关系被集成在统一的关系模型中。

可以将关系和约束两个概念结合起来,用约束对关系进行限制,以使其达到特定设计之要求(图 42.3-13)。为了保证产品快速、高质地投放市场,有必要将产品生命周期中各个业务过程,用某种形式规范化地描述出来,形成跨平台的约束模型。当某一过程考虑对产品的改变时,便自动地接受上游及下游其他业务过程的检验,以确保满足工艺、工装以及制造等各方面的要求,如制造资源约束对设计的检验,制造资源约束要考虑的内容有机床、刀夹具以及量具等,看是否有满足设计的现成的制造资源,只有这样,才能真正实现快速响应产品设计。

(2)对象模型

模型是为了理解事物而对事物作出的一种抽象,在构造具体的系统之前,系统设计者为了不同的目的而构造不同的模型,这就是系统建模。为了形式化地描述系统的对象及对象间的关系,我们必须建立适当的

对象模型。

通过分析,我们发现,对象可以分为两类:元素和关系。

元素对象(Element)表征一种对象类型,如值(长度、角度、变量)、点、线、圆、结构单元等,这些对象具有自己的属性和操作,这些对象不直接参与系统的创建工作。表 42.3-2 列举了其中几个元素对象以及它们的属性和方法,例如,对于线段,它的属性就是起点和终点,线段的颜色/线型等,它的方法有显示、计算等。

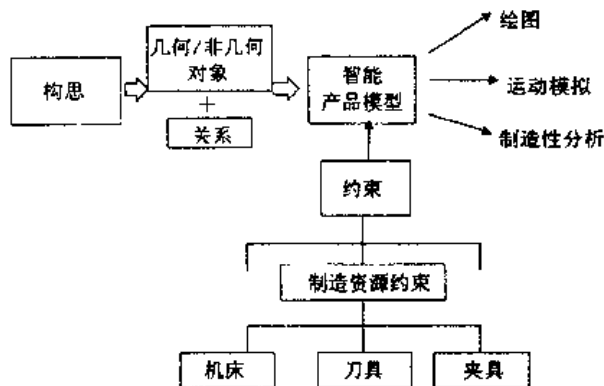


图 42.3-13 关系与约束结合以构造智能产品模型

表 42.3-2 元素类型

对象类型 Type	属性 Attribute	操作 Action
值(Value)	x: real	calculate...
点(Point)	x, y: real	draw...
线(Line)	x1, y1, x2, y2: real	draw...
圆(Circle)	x, y, r: real	draw...
结构单元(Component Unit)	Par-list: Line, Circle, ...	draw...

所有的元素,可以分为两类,一类是基准(支持)对象,另一类是被约束(依赖)对象,在给出关系的定义之前,先给出约束的定义:

约束(Constraint)是联系基准元素和被约束元素之间的纽带。有了约束后,再根据基准元素,即可求出被约束元素。

根据以上定义,我们可以发现,一个对象既可以是基准对象,也可以是被约束对象。约束可以分为两种情况:

- 单重约束:一个对象只受另一对象的约束。
- 多重约束:一个对象受两个或两个以上对象的约束。

约束是关系中的一个非常重要的概念。

据此可定义关系对象(Relation Object)如下:根据已有的对象,受这些对象的约束而建立的对象,是系统通过操作所出现的对象。

一个关系对象类模型主要包括以下一些元素:

- 结果对象(Result Object):一个创建行为的结果,如图 42.3-11 中的切线,用来标识设计对象。

●影响对象(Effect Object):在结果对象之外伴生的对象,如图 42.3-11 中的切点,用来表示对象的作用结果。

●对象行为(Action):保存在系统数据结构中的为创建结果对象而进行的活动,表示对象创建方法。

●对象构造参数(Parameter):创建结果对象所需的约束参数,包括对象参数、值参数、枚举值参数等。

●属性(Property):与对象的创建无关的性质,例如线段的颜色、线型等,用来描述设计对象的功能和结构特征。它们与对象相关,而与行为无关。

图 42.3-14 表示关系对象类模型及其元素。一个关系对象的建立过程如下:

- 1)记录对象类型;
- 2)记录对象的创建过程;
- 3)建立对象与它的参数之间的关系。

最后,可定义关系(Relationship)如下:通过创建行为在结果对象和参数对象之间建立的联系。

对象间的关系,是以创建参数为媒介,通过对象的创建行为来实现的。在图 42.3-11 所示的例子中,各个

对象的创建过程如图 42.3-15 所示:

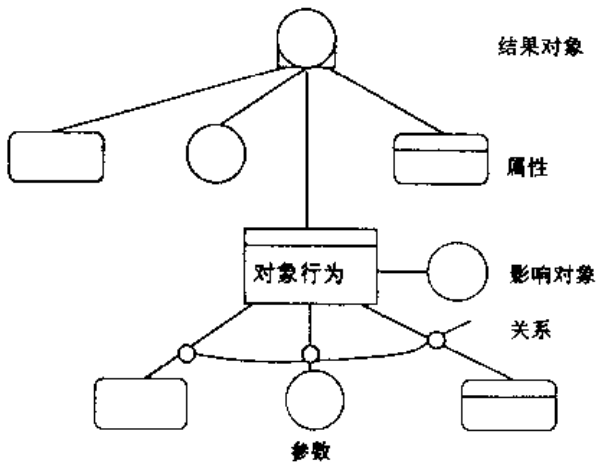


图 42.3-14 关系对象类模型及其描述

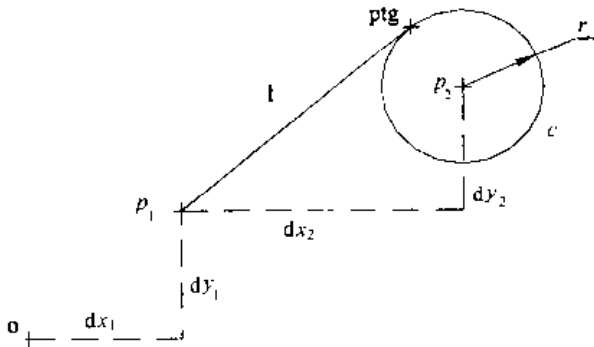


图 42.3-15 图 42.3-11 所示的例子

其中,  $p_1$  以坐标原点  $o$  为参考点, 偏移  $dx_1, dy_1$  建立(图 42.3-16);  $p_2$  的建立过程与  $p_1$  类似, 不过  $p_2$  以  $p_1$  点为参考点, 而非以  $o$  为参考点(图 42.3-17);  $c$  以  $p_2$  为圆心, 以  $r$  为半径, 通过圆心半径方式创建(图 42.3-18);  $p_1$  与  $c$  是直线  $l$  的两个参数,  $l$  通过过一点与圆相切方法建立(图 42.3-19), 切点  $ptg$  是影响对象。在这个例子中, 任何约束对象发生改变, 则约束自动进行传播, 以维持系统中定义的关系不变(参考图 42.3-12)。

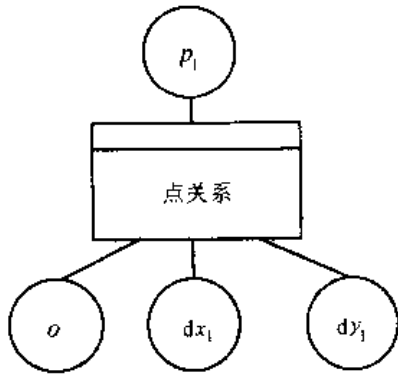


图 42.3-16  $p_1$  的创建过程

(3) 结构单元

在一些传统的 CAD 系统中, 元素的集合称为组、

图块或图段, 这些概念, 或者通过编程实现, 或者只有插入点、旋转角度、尺寸几个参数。在关系型 CAD 系统中, 作为几何对象集合的结构单元, 不仅在逻辑上是一个统一的整体, 而且组成对象间具有几何或数值关系, 还可以建立与单元外部的对象的关系。单元中的任意对象都可以被看作是一个参数, 不只是尺寸和点, 甚至结构单元也可以作为一个参数。

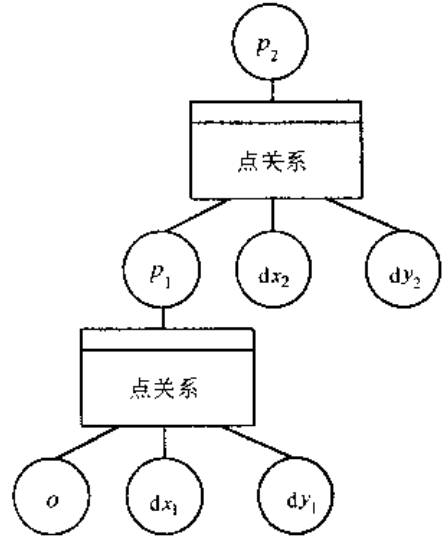


图 42.3-17  $p_2$  的创建过程

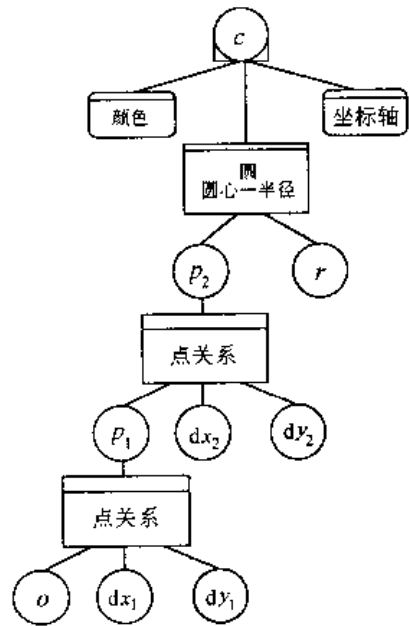


图 42.3-18  $c$  的创建过程

在变型设计过程中, 一个具有相同或相似结构、尺寸不同的零部件族是必需的, 结构单元用来实现这一目标。结构单元分为两个相关的方面: 单元模式(Component Unit Scheme)与单元实例(Component Unit Instance)。模式用来描述单元的内部结构, 它决定单元的组成元素的数目及元素间的相互关系。

单元的一个显著特点是它可以具有参数, 也有两

种类型的单元参数:形参(Formal Parameters)和实参(Actual Parameters)。形参存在于单元模式中,它决定了实参的类型和顺序。实参不属于单元实例,但是被实例用来决定其形状、尺寸和位置。实参定义了结构单元与模型中其他对象间的关系。单元模式可以和多个单元实例联系在一起。当每个单元的实参不同而内部结构一样时会出现这种情况。

结构单元的图形表示称为单元实例,因为从而面向对象方法学观点来看它们可以看作单元模式的实例。在面向对象的实现策略中,我们把这两个概念作为两个单独的类并建立它们之间的关系。

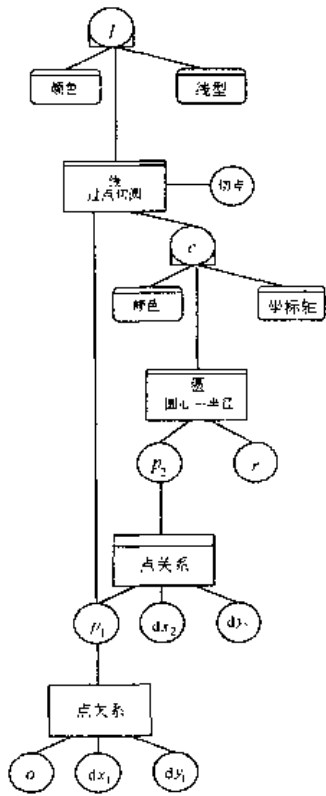


图 42.3-19 I 的创建过程

单元模式包括:

- 形式参数列表;
- 元素列表。

元素对象大都互相关联,其中一些与形参有关。

单元实例包括:

- 实参列表;
- 相应的单元模式。

单元模式类的一个实例是一个以确定数目几何对象为组成元素和一定数目形参的单元模式。一个实在的单元模式不在几何模型中出现,单元实例类的一个实例是几何模型中具体单元模式的表示。图 42.3-20 表示了单元模式类、单元实例类以及它们的实例。

如前所述,单元参数包括形参和实参。其中,形参包括:

- 基本类型,例如,值、点、线以及圆等。
- 单元实例。

形参是单元模式中一些元素对象的支持对象,它们决定了这些元素对象的图形数据。实参必须与相应的形参具有同样的数据类型或派生类型,因此,一个实参可以是一个派生对象。实参不属于单元实例,它们是几何模型中的局部几何对象,如果形参是一个单元实例,则实参必须是同一单元模式的一个实例。

实参的数据通过形参决定了单元实例的形状、尺寸和位置。单元实例的计算过程执行如下的操作:

第一步:将实参的图形数据复制到相关单元模式的形参中去;

第二步:运行相关单元模式中元素对象的所有计算过程来为每个元素生成图形数据;

第三步:返回第二步中得到的所有图形数据给单元实例。

图 42.3-21 展示了实参、形参和单元模式。单元模式中的箭头显示元素对象的支持及依赖关系,不同的实参值产生单元模式的不同变型。当一个单元实例用一个模型的局部对象为实参,它就完全嵌入几何模型中去了,就与模型中的其他对象建立了关系。

形参在结构单元建立时进行定义,然后对应形参的实参可以用事物特性表来表达。

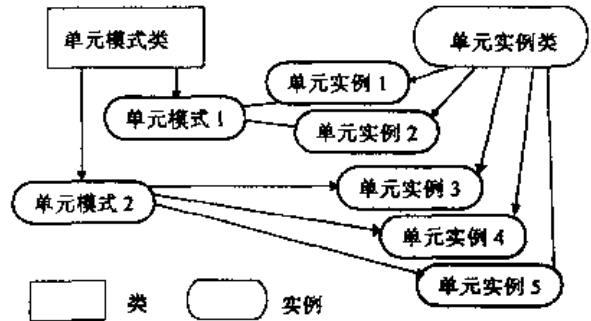


图 42.3-20 单元模式类、实例类及其实例

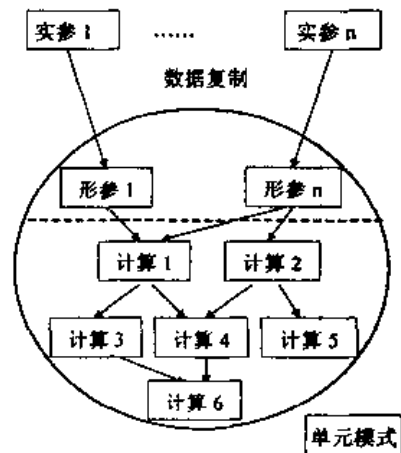


图 42.3-21 单元实例的求解过程

## 第4章 关系型产品模型理论及其应用

### 1 关系型产品模型的定义及应用背景

产品模型是一种数字化的信息模型,它以一定的数据抽象,定义和表达在产品生命周期中有关的信息,包括数据、结构和关系等。产品模型作为表达产品数据及其联系的信息模型,与它所描述的对象——产品活动过程有着密切的关系。传统 CAx 环境下的产品活动过程,如 CAD, CAPP, PPC 等,虽然也应用计算机,但是它们各自有自己的数据属性、数据格式和存储结构,从而使得产品活动进程出现中断,延缓了进程速度。同时还增加了数据冗余,数据一致性维护要求很高。而在基于产品模型的产品活动过程中,引入了产品模型,使得数据描述的准确性、一致性、完备性得到保证,数据不再冗余,产品活动过程可以畅通无阻,从而加快了活动进程。

目前的产品模型主要有以下几种:基于特征的产品模型,面向装配的产品模型,面向功能的产品模型,基于 STEP 的产品和关系型产品模型。其中关系型产品模型正是为重用产品信息的目的而提出的。

所谓关系型产品模型 (Relation-based Product Model, RPM),它是根据合理化工程哲理的订单规划的需要而提出的。它是在先进的信息理论和计算机技术的支持下,为重用产品资源而定义和表达的产品数据及其数据关系的数字化的信息模型。在产品模型中,显式地描述产品开发最终结果的数据内容简称产品数据;而在产品的数据定义过程中,隐式地定义了产品数据与活动过程的关系以及产品数据值之间的联系,这些统称为数据关系。分析表明,实现订单产品的快速、合理的变型是由这些隐式关系决定的。因此,将产品数据和在产品开发过程中定义的种种数据关系,统一地表达在数字化信息模型中,就成为关系型产品模型。图 42.4-1 表示关系型产品模型的原理和它在订单规划过程中的作用。图 42.4-1a 中,将产品模型的信息结构从逻辑上分为四个子模型,它们共有的内核是关系和事物特性,其中,关系是于模型内部及子模型之间数据关系的抽象;而事物特性则定义与表达各个子模型的产品数据。基于关系型产品模型的支持进行用户订单的变型规划时,在产品项目小组的管理和控制下,在图 42.4-1b 中,工程师们根据定义在产品数据管理 (Product Data Management PDM) 系统中的重组后业务流程,利用关系型产品模型中的相关子模型和工程

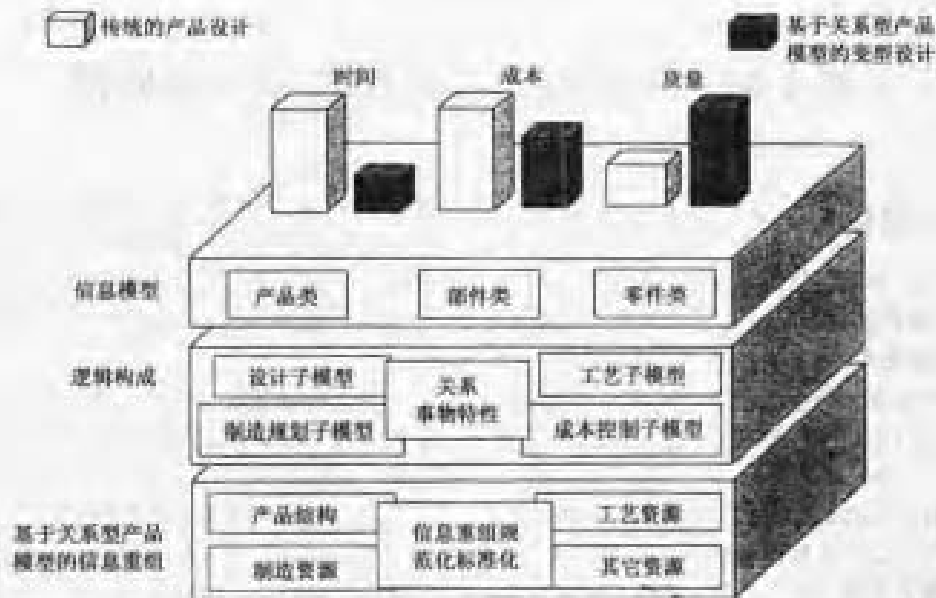
数据库,并且在制造资源计划 (MRPII Material Requirement Plan II) 系统的支持下,并行地对用户需求作出合理的变型规划。目标是当天作出有关交货期和价格的准确响应,并从当天起组织变型产品的制造。

本章着重介绍其中设计子模型对产品数据和数据关系的定义和表达。在设计子模型中,基于面向对象的建模技术,将产品信息资源分为产品类、部件类和零件类三大对象类,并将它们分别又按信息类型用分类定义模型 (CDM)、形状定义模型 (FDM)、事物特性定义模型 (ACDM)、视图定义模型 (VDM) 和信息管理模型 (MM) 等加以定义和表达 (见图 42.4-1a)。这些模型合起来统称对象类的类属模型,能很好地描述一个对象的分类、形状、参数、视图和管理等有关的属性定义和操作定义。

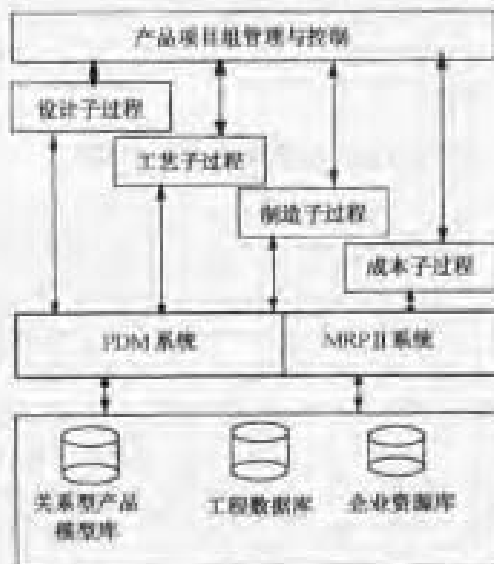
#### 1.1 变型设计在订单规划中的作用

前已说过,市场竞争的全球化,使得用户有了更多的选择对象。用户需求的多样化、个性化,导致多品种、小批量 (乃至单件) 的订单方式在许多机械制造企业的普遍实行。与传统的少品种、大批量生产方式相比较,在订单式生产中,由于品种的增多和批量的减少,使得产品信息类型更加多样,数量更加庞大,因而造成空前复杂的局面。但在另一方面,企业要在竞争中取胜,又必须以比同行更快的交货期、更低的成本、更好的质量和更完善的服务去响应用户的需求。要达到此目的,企业唯有科学有序地利用现有的产品信息资源,以变型设计快速、合理地响应用户订单。

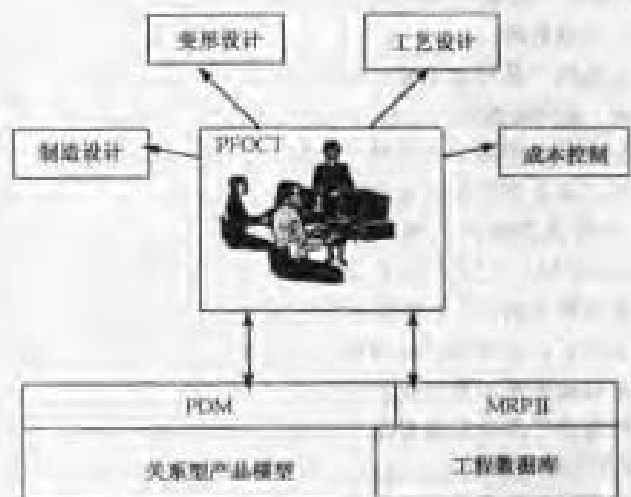
如第 1 章所述,机械产品的设计,通常可分为新颖性/新型设计和适应性/变型设计两大类。新型设计活动是一个从无到有、从顶向下、从粗到细的信息转换过程。它从产品需求开始,首先通过抽象,定义一个功能空间 (Function Space, FS); 然后运用有关知识,确定一个适应于 FS 的解决方案,也称为行为空间 (Behavioral Space, BS); 再经过分析计算,构造对应于 BS 的结构空间 (Structure Space, SS)。变型设计的活动过程,虽然也存在 FS、BS 和 SS,但与新型设计不同的是,它乃自底向上重用 (或者局部改变而非新造) 这三个已经存在的语义空间的信息,去寻求能满足用户需求的解。设计者首先查找 SS 中是否有能满足功能需求的产品结构;若没有,就访问 BS;再通过适当



a) 关系型产品模型的组成体系



b)



c)

图 42-4-1 关系型产品模型的原理和在定单规划过程中的作用

修改 BS 中相似的解决办法，以重组 SS 中的产品结构。可见，变型设计能使产品具有很强的灵活性和适应性，能够满足不同用户的多方面需求，因而特别适合于多品种、小批量的竞争性产品生产方式。

在机械制造业中，新型设计毕竟稀少，大量的总是变型设计。事实上，许多设计人员的确是经常使用变型设计去响应用户订单，但是这种传统的变型设计并不能自动保证订单产品的快交货、低成本和高质量。其根本原因在于变型设计的过程中涉及大量繁杂的产品信息和设计知识，而目前这些信息和知识的检索和管理主要仍靠设计人员的手工实现和自发进行，因此严重影响设计的速度和合理性。为了克服这个障碍，就

需要运用先进的信息理论和计算机技术，着手对企业的产品信息资源进行重组。

## 1.2 企业产品信息资源重组

### (1) 产品资源重组的意义

快速响应变型设计以重组企业的产品信息资源为基础。但是，传统的变型设计并不能保证订单生产的快交货、低成本和高质量，有时甚至适得其反。主要表现在企业生产零件的多样性和缺乏可重用性，并导致新零件数目的失控。企业中许多原有零件的功能相同、结构相同、形状相似，但就是不能划一，它们之间的尺寸差异有时相差不过零点几毫米。这种多样性必然引起

制造和管理的复杂化与延滞,以及库存量的增加,与此同时,这些零件的可重用性往往也差,设计需用时难于查寻,有时即使找到了也未必能配上,这种情况容易造成新零件数目的失控。许多新的零件被源源不断地设计出来,紧接而来的就需要动用过多的人力、物力资源去支持后续的工艺、工装、试制和试验等过程。据统计,即使新零件只占到零件总数的10%不到,它却需要占到产品总成本的70%以上,可见减少新零件数目的重要性,引起这种失控现象的原因,在于企业产品信息资源的无序状态和设计者的随意性。在传统的企业中,原有的产品信息资源,通常若不是放在设计者的头脑中,记在他们的本子上,就是存在于资料室的档案内,往往平时泛览,用时却无着落,加上对设计者缺乏必要的约束机制,因而形成新零件的不合理出现,严重削弱企业的综合竞争能力。要使无序的企业产品信息资源变成能有效支持快速响应变型设计的宝贵信息资源,就需要对它们进行规范化、标准化重组。

### (2) 产品信息资源重组的原则

进行企业产品信息资源重组,应遵循以下原则:

1) 强调全局优化 设计一个产品,不仅应充分满足对它的功能需求,还须仔细考虑在实现它的制造、装配过程中的合理性。

2) 坚持规范化、标准化 调查表明,设计的随意性是零件多样化的主要原因,故须以一定的规范和标准进行约束,方可有效减少新零件数。

3) 旨在重用 重组的目的在于重用,故需要对企业的产品——进行分析、归类、优选、抽象和定义,方能保证重组后的产品信息资源具有良好的可重用性。

4) 实现信息化 企业产品资源经过层层抽象,必须表达在一定的信息环境之中,以充分发挥计算机的威力,这就要求在重组过程中运用先进的信息理论,实现产品资源的信息化。

### (3) 产品信息资源重组的内容

实施企业产品信息资源重组,要从企业所有的产品构成出发,以各类相同或相似的产品构成为对象,面向产品活动的全过程,以规范化、标准化为准则,实现对产品构成定义的信息化抽象。在本章中,基于面向对象建模技术,将同类产品构成抽象定义为对象和对象类。在逻辑上,可将产品资源重组划分为功能重组、结构重组、命名重组、参数重组和视图重组等几个构成流程。图42-4-2基于IDEPx原理,表示产品信息资源重组的构成语义。其中:功能重组(Function Reengineering)是关于对象类的功能定义;结构重组(Structure Reengineering)是关于对象类的结构定义的再设计,重组后的结构树(Structure Tree, ST)是对象类的形状构成的定义;命名重组(Naming Reengineering)是对

于对象类的命名定义的再设计;数据字典(Data Dictionary)则是重组后对象类的命名定义;参数重组(Parameter Reengineering)是关于对象类的非形状信息定义的再设计;重组后的事物特性定义模型(Article Characteristics Definition Model)是关于对象类的参数属性的定义;视图重组(View Reengineering)是关于对象类的视图表达定义的再设计,重组后的视图定义模型(View Definition Model)是该对象类的视图表达的原型定义。以上各个构成对于企业产品定义的规范化、标准化的影响侧面各不相同,它们的综合作用才是对企业的产品定义的完整、彻底的再设计。

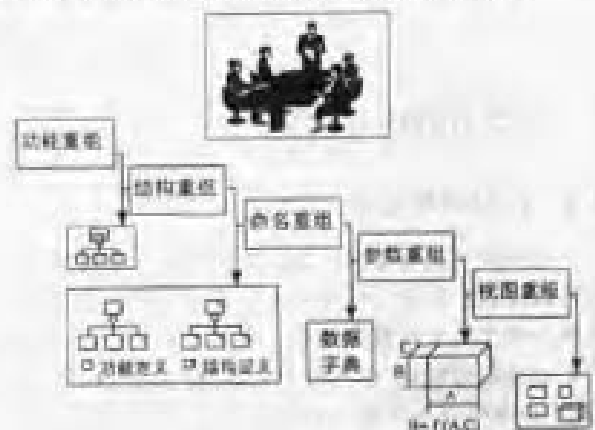


图 42-4-2 产品重组流程

在整个重组过程中,功能重组和结构重组是实现产品定义规范化、标准化的关键。下面用一个实例,简单说明它们的重组过程和内容。图42-4-3表示桥式起重机中一类主要矩形肋板的结构重组实例。这一类零件对象的功能是增强主要结构纵剖面的局部刚度,它与主梁的联接由焊接实现。此外,其他约束条件/功能还包括:对主要结构自重的影响最小,联接焊缝与主要结构自身的焊缝不交叉,沿主梁结构长度方向内部挖孔供布线(电力线)行进。图中a显示企业在这一类零件定义中的形状多样性。重组后的这类零件功能树(FT)如图b所示。其中:功能定义是增强主要结构纵剖面的局部刚度(FT),它的子对象包括构造肋板形状(FT1)和确定与主梁结构的连接(FT2)。FT1的下一层子对象又分,定义截面特性最大,对主要结构自重影响最小的肋板形状(FT11)和定义穿孔形状(FT12)。FT2的子对象则包括:定义与主梁连接的焊缝类型(FT21)和定义与主梁结构自身焊缝不干涉的过孔形状(FT22)。另一方面,重组后这类零件的形状构成定义的结构树如图c所示。图中,这类肋板零件的形状构成(ST)包括四个定义,即基本形状定义(ST1)、穿孔形状定义(ST2)、坡口形状定义(ST3)和过孔形状定义(ST4)。对于每个子定义,又有相关的子定义,例

如 ST3 又有子定义 ST31、ST32、ST33 和 ST34。不过，这一层子定义与上层的语义关系是 a-kind-of 的类属关系，而非上一层的 a-part-of 构成关系。在图 42-4-3 中，图 b 对应于产品设计过程的功能空间 FS，图 c 对

应于行为空间 BS 或概念模型，而图 a 则对应于结构空间 SS。有了这样的产品信息资源重组，规范化、标准化的结构形状定义取图 a 中的左起第一个。

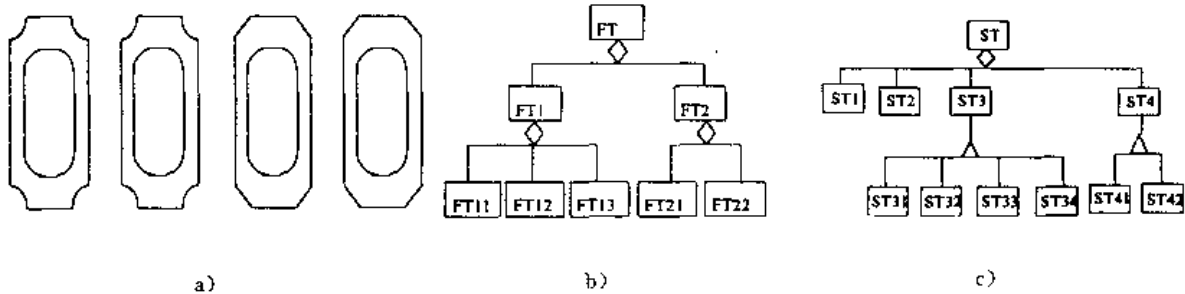


图 42.4-3 一类矩形肋板零件的产品定义

## 2 对象类的信息构成

### 2.1 产品对象定义

关系型产品模型 (RPM) 在快速响应变型设计中的特殊作用，决定了它必须将与有关产品数据的产生及重用密不可分的数据关系作为定义和表达产品定义的不可分割的组成。作为解决可重用产品定义的理论，RPM 必须研究不同类型的产品定义的信息构成，必须研究不同层次的产品定义的信息构成，必须研究统一、完整地表达这些信息的技术实现。因此，一个合适的关于产品数据和数据关系的抽象定义与表达的模式应该是 RPM 的研究的核心之一。在这一节里，将基于对象建模技术 (OMT) 原理，引入产品定义对象 (PDO)，来抽象定义不同类型、不同层次的产品定义，对象 (Object) 和对象类 (Object Class, OC) 是关于这种抽象的统一表达。

#### (1) 对象建模原理

机械产品设计活动的明显特征可概括为：过程复杂，数据类型丰富，数据量大。研究工作表明，传统的 CAD 平台只表达设计活动最终结果，丢失了大量的表达设计意图和设计过程的信息，因而制约着在快速响应变型设计中重用这些结果。工程师们不仅仅需要对最终结果的清晰表达，而且还要更多地关心得到这些结果的 know how (技巧)。正因为这样，在产品模型的研究中正在关注这一部分信息的定义与表达，其中，基于数据抽象的理论来研究产品定义的语义描述是非常重要的内容。

数据抽象是关于应用域内某些细节内容的一个描述模型，数据抽象的目的是为了强调表达那些研究域中所关心的问题的细节，而忽略那些尚不涉及(至少暂时不涉及的)的问题。数据模型则是由一定的数据模式对数据抽象的定义与表达。列表描述所涉及的数据对

象的记录，大概是最古老但至今仍在沿用的一种数据模型。一旦确定了列表的结构，则只需不断地往表中填入相关的记录。伴随信息理论的研究进展，特别是计算机技术的发展而不断发展的数据库理论的研究，数据模型的演化经历了文件模型、层次模型、网状模型、关系模型及面向对象模型。关系数据模型 (Relational Data Model, RDM) 由于其成熟的关系代数的支持，加之它描述数据对象的直观性，使得数据库理论在商业化应用中很广泛，直到目前，关系数据库 (Relational Database, RDB) 仍是商品化数据库产品的主导。RDM 应用主要地集中在相对稳定的数据模式并且有大量重复数据的场合。随着信息理论和计算机技术的研究进展，一些非传统的应用领域，如计算机辅助技术 (CAD、CAM、CAE)，数据的类型及内容都发生了很大的变化。这些应用领域的特征可概括为：

- 结构化数据呈现更多的动态变化，并且常常与重复性数据相互牵连；

- 结构化数据的比重远远小于传统的简单信息内容的场合；

- 最终用户可以对相同的数据有不同的描述。

正是由于这些应用领域问题的特殊性，源自面向对象编程 (Object-Oriented Programming) 的面向对象数据模型 (Object-Oriented Data Model, ODM) 在工程问题的抽象定义与表达的研究中得到广泛应用。下面通过一个产品族对象类的结构定义的数据抽象，具体地比较 RDM 与 ODM 的不同处理策略。

产品族对象类是指具有相同的功能定义、相同的行为定义、相似的结构定义的一类产品对象 (详见 4.2 节)。它们的结构定义描述了组成该对象类的同类对象。图 42-4-4 是一个齿轮轴部件类的结构定义。对于这样的定义，如果采用结构树的通用描述，其层次结构中每个节点的不同对象，就构成了不同的变型实例。在产品定义中，对图 42-4-5 这样的结构树的通用描述，



等同于某应用域行为空间的定义；对它的某个具体可能解的特定描述，等同于功能空间中一个特定需求的变型产品的结构定义。因此，一个合理的数据模型，不仅应能表达通用描述，而且应能表达特定的变型描述。

关系数据模型(RDM)是建立在数学概念基础上，用数学方法来处理数据的。在前人工作的基础上，美国IBM公司的Codd从1970年起连续发表论文，系统地、严格地提出了RDM的理论。在RDM中，无论是实体集还是实体之间的联系都用关系来表示，数据在用户的观点下的逻辑结构是一张二维表，表的结构(或这种关系)实际上就是一个特定数据模式。不难看出，这样的表达在数据冗余方面存在有明显的不足。这种冗余，一方面表现在不同抽象层次的数据模式的定义中(如产品族构成定义记录与产品某个变型实例的记

录)；另一方面也存在于同一抽象层次的数据模式的定义中(如产品变型的零件记录中将会有许多冗余记录)。这种冗余带来的问题在于最终实现这样的数据库将会因此而庞大、低效。RDM在表达工程问题的抽象时，另一个主要的问题是关于过程与数据的分离，这给数据的重用带来很大的困难。图42.4-5a是基于RDM的图42.4-4的结构定义的形式化的表达。

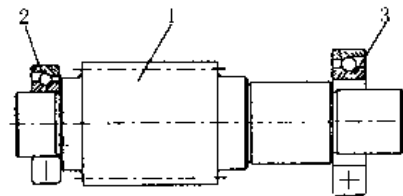
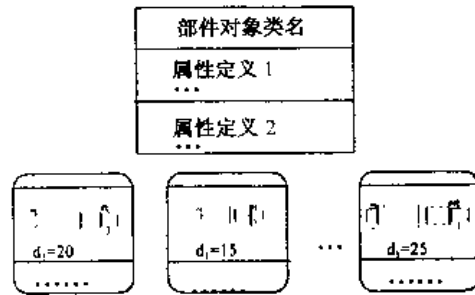


图 42.4-4 一个齿轮轴部件类的结构定义  
1—齿轮轴 2—左轴承 3—右轴承

识别号	齿轮轴	左轴承	右轴承
458579	552522	362541	568724
654584	313564	478541	654125
.....	.....	.....	.....

a)



b)

图 42.4-5 图 42.4-4 的数据抽象

面向对象数据模型(ODM)按照真实世界的对象和概念来构造模型以表达信息，它的基本单元是对象(Object)，它包括了数据的结构和数据的行为，是一种抽象的定义。它可以是一个物理的对象(如齿轮轴部件类的结构)，也可以是一个抽象的描述(如齿轮轴部件类)，还可以是一个概念(如运动转换)。在ODM中，对象以属性和操作作为自身标识，对象类(Object Class)定义了一类具有相同的属性定义和操作定义的对象，对象类也简称类(Class)，其子类(Subclass)对类具有继承性，即：子类继承有类的全部特性、行为以及与其他对象的关系。子类仍然可以有自己独特的属性定义与操作定义。此外，在ODM中，还可将对象的行为(或关于对象的操作)封装在对象中。这样，在访问对象属性时可以直接知道对象的活动过程。这在工程问题的建模分析中特别重要(详见下一节的分析)。ODM的重要特性，除上述继承性外，还有标识唯一性、分类性、多态性。图4.14-5b是基于ODM的图4.14-4的结构定义的形式化的表达。

数据模型是客观世界抽象结果的表达模式。RDM是商用数据库研究的理论基础，但它对工程问题的复

杂对象的表达缺少有效的支持。ODM则为表达工程问题的复杂对象提供了新的理论。不过，由于它仍是在发展中的理论，因而尚无完善的、通用的理论与方法支持。在本章的研究工作中，综合RDM与ODM，提出了产品定义对象(PDO)的理论框架，实现对于规范化、标准化的产品资源的抽象定义，是关系型产品模型中定义与表达不同层次的产品资源的理论基础和基本出发点。

(2) 产品定义对象

产品定义的数据与数据关系是可重用资源的重要构成，因此，它们是关系型产品模型(RPM)理论研究的核心。基于面向对象数据模型(ODM)的原理，将不同层次的产品定义加以抽象，用统一、完整的抽象定义来表达它们的属性与操作。产品定义对象(PDO)正是关于这种抽象的理论。

产品定义对象(Product Definition Object, PDO)是关于不同层次的产品构成本身的数据和数据关系的抽象，对象和对象类的属性定义和操作定义是这种抽象的定义与表达。

作为不同层次的产品构成，可以是零件对象，可以

是部件对象,也可以是产品对象。虽然,它们在新产品开发循环中定义的数据和数据关系的类型与内容是各不相同的,但是,作为机械产品共性定义—功能定义、结构定义、形状定义、非形状定义、视图定义等,在它们的抽象定义中应该是共同的。在 ODM 中,对象与对象类的个体与一般的关系由继承性保证,因此,PDO 将这些不同的构成抽象为对象和对象类,每个对象类的属性定义和操作定义是这一类对象所共有的。还有,基于 ODM 原理,将对象的属性定义与操作定义封装在对象类的定义中,对象在不同重用场合所需的重用定义,可以由被封装的操作定义决定的不同操作和属性定义决定的不同取值来实现, PDO 的语义表达如图 42-4-6 所示。

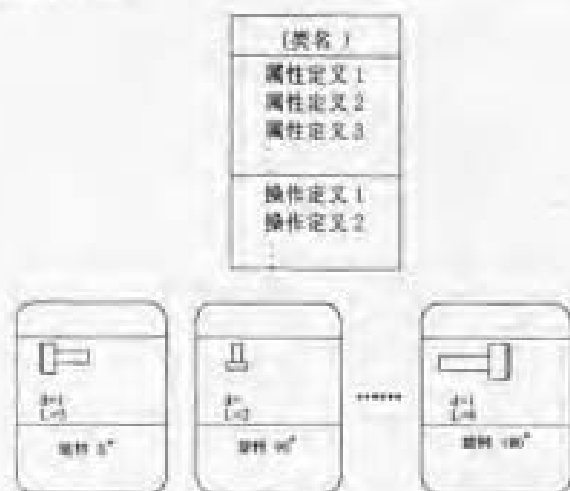


图 42-4-6 PDO 的语义表达

在 PDO 的框架下,对象和对象类是不同层次的产品构成本身的抽象。其中,对象(Object)是在 PDO 的框架下关于产品构成本身的抽象,属性和操作是它的标识。对象类(Object Class)是关于一类具有相同的属性定义和相同的操作定义的对象抽象。

引入对象和对象类,使得基于属性定义和操作定义来描述一类产品构成的数据和数据关系成为可能。在本章的后续部分将就对象和对象类展开讨论。由于本文工作的重点在研究信息环境下订单规划变型设计的合理、快速的实现,因此,在 PDO 的框架中将产品构成本身的抽象定义包含定义状态和重用状态是必要的。

一般说来,对象类在被规范化、标准化地抽象定义为 RPM 时,它的形状定义、非形状定义、视图定义等等处于定义状态。当它的一个对象被用于一个给定的应用,如被用于与它类对象装配,被用于构造工作图原型等,这时的定义属于重用状态。一个法兰零件对象类的定义状态与重用状态的示例如图 42-4-7 所示,图 a 是该类的一个定义状态,图 b 是参与装配的该类重用状态。在 RPM 的研究中,对象类的定义状态与重用

状态是关于该类的一个完整定义,通过对象类的必要的属性定义和操作定义被封装于对象类的抽象定义中。

在研究产品定义的对象类的过程中,有时需要描述从功能空间的定义去检索结构空间的结构定义,有时需要在定义过程中对于对象类有直观的结构描述,有时也需要在定义过程中用对象类的结构定义进行过程描述。为了方便地实现这些应用需要,有必要定义结构单元(Component Unit, CU)的概念。

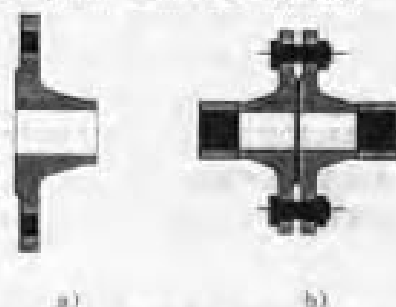


图 42-4-7 对象类的定义状态与重用状态示例

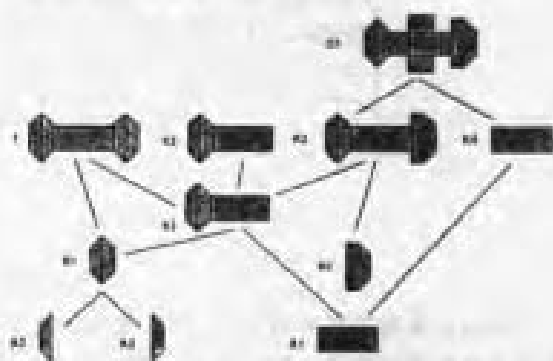


图 42-4-8 基于 CU 的产品构成关系

结构单元是关于产品构成本身的一个抽象的语义表达。在不同层次的产品构成中, CU 可以表达不同的对象类。例如,在 RPM 理论中,零件对象类的结构定义是由零件对象结构重组后的规范化、标准化的单元形状来表达的,因此,零件对象类的 CU 就是这些单元形状。又例如,部件对象类的结构定义是由相关的零件对象类(或零件对象类与子部件对象类)的形状定义来表达的,那么,部件对象类的 CU 就是这些相关的对象类。正如 PDO 的引入实现了关于不同层次的产品构成本身的的数据与数据关系的抽象, CU 的引入在表达一个对象类结构定义时提供了很好的抽象。图 42-4-8 描述了基于 CU 的不同层次的产品构成关系。

形状的定义与表达是机械产品信息模型的一个重要方面,自然,这也是 RPM 的研究应予以关注的。形状属性作为对象类的形状定义的抽象,包括了定义状态与重用状态的定义。那么,形状定义的过程中应该定义哪些信息才能达到这样的要求?在讨论之前,先来研

究图 42.4-9 所示曲柄-滑块机构的行为空间定义的概念模型的建模过程。图中 a 是该机构以形状定义表达的概念模型，图中的滑块已被抽象为一个点  $P_2$ 。对于这样一个概念模型，它提供了哪些信息呢？首先是该模型的几何数据描述，易知这样一个概念模型是由三条一定长度的线段 ( $L_1, L_2, L_3$ ) 和三个点 ( $P_0, P_1, P_2$ ) 构成。那么，是不是由任意的三条线段和三个点就能表达所需要的行为空间的定义呢？或者说，由任意三条线段和三个点能表达“图解法”所表达的有定义呢？答案是否定的！这是因为，由任意的三条线段和三个点所构造的“形状”在本质上只是几何元素的定义，对于这样的“形状”的理解是由设计人员的智能解释决定的。我们知道，一个形状的正确定义由几何元素和拓扑关系构成，只有将这样的关系定义作为表达形状定义的信息模型的组成，才是完整地定义了对象类的形状属性。在本例中，图解法的定义过程在这些几何数据

之间确定了相应的拓扑关系，即图中隐含的角度关系 ( $\angle P_1P_0P_2, \angle P_0P_1P_2, \angle P_0P_2P_1$ )，正是由于它们的存在，才确定了图示的形状定义！于是，有图 42.4-9b 所示的 a 中构成对象的语义表达。问题是，在信息环境下如何定义这样的关系信息？还有，对于更加复杂的构成对象的关系信息，如部件对象中零件对象之间的装配关系，又该如何定义？我们还想知道，如果修改图中的这些几何数据和它们之间的拓扑关系的任何一个定义，结果又会是怎样呢？当然，作为一个信息模型，它还应该提供决定有关结构空间的参数的“源”。换句话说，它应能提供确定有关结构参数的关系。例如，这样的模型应该能表达运动仿真信息，并且可将这些信息被结构设计所继承，它也能分析动力学运算结果（如在一定的物理结构下面作角速度、角加速度分析）等等。所有这些需求，决不能只是对数据描述的抽象就能办得到的，它需要对数据关系抽象的支持。

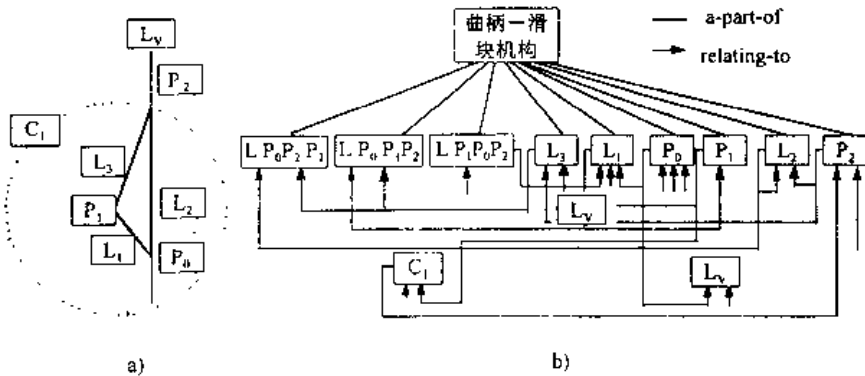


图 42.4-9 曲柄-滑块机构的概念模型

将数据与行为归并在对象的定义中，是对象建模技术 (OMT) 的基本要素之一。换句话说，无论是对象还是对象类都可以包含有关的操作 [Operation]。这里的操作可以是作用于一个类中的对象的某个功能或变换，也可以是一个类的对象所施加的某个功能或变换。在机械产品的设计中，形状定义是非常重要的，它与很多的过程操作有关，如几何元素的定义、几何元素的关系定义等。同时，在这些过程操作中，还会有很多的取值关系的定义，如齿轮分度圆直径与齿轮的齿数与模数的关系等。基于 OMT 的基本原理，在产品定义对象 (PDO) 的框架下，考虑到机械产品设计中不同层次对象类的形状定义中数据及隐含的过程和关系的复杂、多样，为了不失一般性，可以引入元对象模型的概念。元对象模型 (Meta Object Model, MOM) 是基于 PDO 定义框架，关于对象类的形状定义所拥有的数据和数据关系的抽象，属性、行为、参数、影响以及显示特性是它的标识。其中，属性 (Attributes) 是元对象的数据描述。行为 (Action) 是元对象属性定义的话

动过程。参数 (Parameters) 是元对象的行为定义中涉及的输入参数。影响 (Effects) 是元对象的某个 (或某些) 属性对它对象施加的作用。显示特性 (View Properties) 是元对象在视图定义中需要的显示属性的定义。MOM 的语义描述如图 42.4-10。

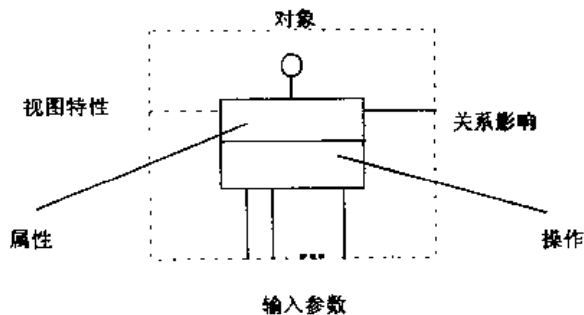


图 42.4-10 元对象模型

上述定义表明，元对象模型具有如下这样一些特征：

- 抽象性 与 OMT 中的对象一样，这里的元对

象定义既可以用于行为空间概念定义的抽象,也可以用于结构空间的形状定义的抽象。这种抽象性的意义在于本章所提出的 MOM 可用于定义与表达规范化、标准化的产品资源的形状定义。

●层次性 在一个特定的产品定义抽象的语义模型中, MOM 可以表达不同层次对象类的形状定义。在产品层,它可以定义一个产品对象类的形状定义。在零件层,它可以定义一个零件对象类的形状定义,等等。

●完备性 在 MOM 的抽象定义中,将对象类形状定义的数据和行为归并在一起,这样得到的语义模型,在表达对象类的形状定义信息方面是完备的。

●一致性 机械产品的不同层次的形状定义,包含了不同的数据与数据关系。基于 MOM 的抽象定义,这些数据与数据关系的抽象属性是相同的,因此,不同层次的形状定义在被抽象后总可以用上述定义加以一致地表达。

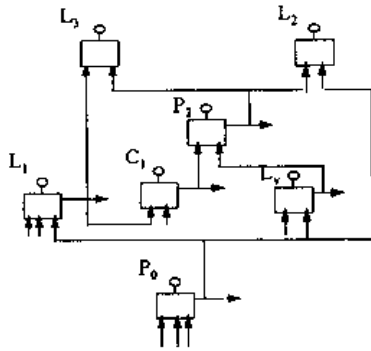


图 42.4-11 基于元对象模型的  
图 42.4-10 的形状定义

MOM 的引入,使得在 RPM 的理论中用统一的面向对象的数据模型来定义与表达不同层次产品的形状定义成为现实。仍以图 42.4-9a 中的曲柄-滑块机构为例,说明 MOM 在构造行为空间的概念模型中的应用。由“机械原理”基础知识,曲柄 L<sub>1</sub> 是以 P<sub>0</sub> 点为始点的有向线段, P<sub>2</sub> 点是滑块沿垂直方向运动的轨迹,同时, P<sub>2</sub> 点又是连杆 L<sub>3</sub> 上的点,连杆 L<sub>3</sub> 除了 P<sub>2</sub> 一个端点外,另一个端点 P<sub>1</sub> 也是曲柄 L<sub>1</sub> 的终点。因此,确定 P<sub>2</sub> 的正确操作是过 P<sub>1</sub> 点做辅助圆 C<sub>1</sub>, C<sub>1</sub> 与垂直辅助线 L<sub>2</sub> 相交即得点 P<sub>2</sub>。最后,连接 P<sub>1</sub>、P<sub>2</sub> 得 L<sub>3</sub>, 连接 P<sub>0</sub>、P<sub>2</sub> 得 L<sub>2</sub>。这样得到的图 42.4-9a 的概念模型,应用 MOM 抽象定义后,即可得到图 42.4-11 所示的曲柄-滑块连杆机构的形状定义的语义表达。

在关系产品模型 (RPM) 的理论体系中,产品定义对象 (PDO) 是关于不同层次的产品构成本身的抽象定义,对象和对象类的属性定义和操作定义是它的数据与数据关系的抽象定义。基于 PDO 框架,结构单

元 (CU) 是关于产品构成本身在结构定义的表达中的抽象,由 CU 可以统一地描述不同层次的对象类的结构组成。形状定义是机械产品设计中必须的,形状定义的完整性要求对象类的抽象定义不仅适用于定义状态,而且也要适用于重用状态。元对象模型 (MOM) 的引入,使得完整定义与表达对象类的形状定义所必需的数据与数据关系成为现实。基于 PDO 框架,对象和对象类具有的高度抽象性,为在 RPM 的理论与方法的研究中,统一、完整、准确地定义与表达规范化、标准化的产品资源提供了理论基础。

### 2.2 各产品对象类的信息构成

如上所述,在产品模型中,包含显式和隐式两类信息。显式信息描述产品的最终结果,对应于产品、部件和零件等物化存在的有形表达。常用的表达形式是工程图样、工艺卡片、制造文档等。在这些有形表达的规划过程中,在它们的物化过程中,则需要知识和规则,需要对各种数据进行提取和加工。这些大量存在于有形表达的规划过程和物化过程的信息,称为隐式信息。数据关系就是这些隐式信息的抽象和归纳。产品模型中隐含的关系,从所定义的过程划分,可分为过程内部关系和过程之间的关系。例如,在产品开发过程中,存在着许多算术运算式,存在着许多逻辑表达式,存在着许多函数关系式,这些属于形式化的关系;与此同时又存在着功能关系、装配关系、构成关系等,这些属于非形式化的关系。此外,在概念模型、装配模型和零件模型的内部和它们之间,还存在着对精度属性、可制造性、可装配性和成本控制的思想等,这些也是非形式化的关系。图 42.4-12 是部件对象类信息构成的语义表达。

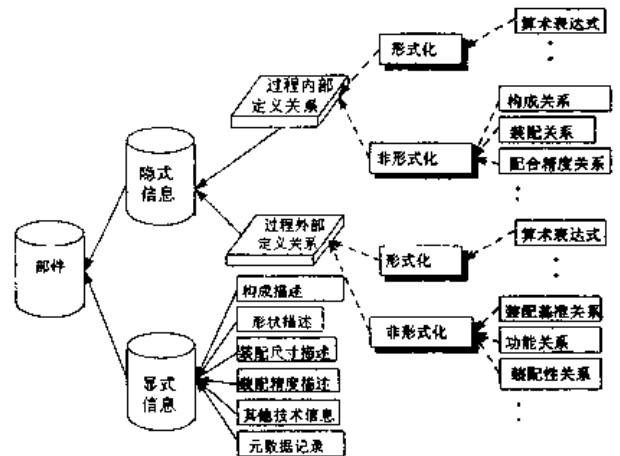


图 42.4-12 部件对象类信息构成的语义表达

### 3 对象分类框架

在企业的大量产品信息资源中,存在着许多不同

层次、不同类型和不同表达方式的数据定义。因此，一个合理的对象分类框架，是有序管理这些资源的基本保证，也是关系型产品模型的重要组成部分。

随着企业产品资源的丰富，有效地、有序地管理这些资源的定义与重用是非常重要的。否则，无序的信息将导致不良的决策，或导致决策响应慢，并且由于难以查找已有的产品资源而导致新的对象类或对象产生，或导致数据错误或产生附加的数据，使得成本上升。计算机技术的广泛应用，使得数据处理的速度与精度大大地提高。但是，它本身并不能解决上述这样重要的问题。产品定义对象 (PDO) 是关于产品本身的抽象，对象和对象类是这种抽象的定义与表达。在企业的产品资源的定义中，大量存在着不同的层次、不同的类型及不同的表达的数据定义。因此，一个合理的分类这些资源的理论与方法，是 PDO 得以成功应用的保证，当然，也是关系型产品模型 (RPM) 理论体系的重要构成。在本节中，基于 PDO 的定义框架，汲取成组技术 (GT) 的分类原理，研究面向企业不同层次的产品资源的统一的并具有唯一性和完备性的对象分类框架 (OCF) 的理论与方法。

### 3.1 GT 分类原理

成组技术 (Group Technology, GT) 是一个哲理，它通过强调产品及其生产活动中隐含的相似性，帮助管理制造和工程活动中的多样性。分类 (Classification) 及编码 (Coding) (C&C) 作为 GT 重要的基本理论之一，其研究已经趋于成熟。根据分类所采用的编码结构，可将 C&C 分为三种模式，它们是：单段码 (monocode)、多段码 (polycode) 和复合码 (hybrid)。单段码的分类模型是一个层次树状结构，它的任何一个中间节点描述了一个特性，可按照由特殊到一般的层次加以展开。如零件的形状特性可以划分为旋转体和非旋转体，旋转体的形状特性又依次可划分为定中心的、同心的、齿轮轴等等。这种模型的明显特征可描述为每个节点的属性是排他的或互斥的。多段码的分类模型借助预定义的字段定义，将一些看似非层次关系的相关特性包括在同一分类定义中，如几何结构、材料、初始状态等。这种模型的每个字段表达一个预定义的特性分类，因此，它是相容的。复合码的分类模型实际上是由单段码与多段码复合而成，它汲取了这两种模型各自的优点，能够完整地分类制造过程所关心的零件分类定义。基于对象建模 (OOM) 的原理，将 OOM 中对象和对象类的概念用于指导 C&C 的设计，基于归纳 (Generalization)、聚合 (Aggregation)、关联 (Association)、分类 (Classification)、作用名 (Role Name) 等语义抽象，扩

展了传统的 C&C 系统。图 42.4-13 是基于聚合 (Aggregation) 语义抽象对零件的分类示例。

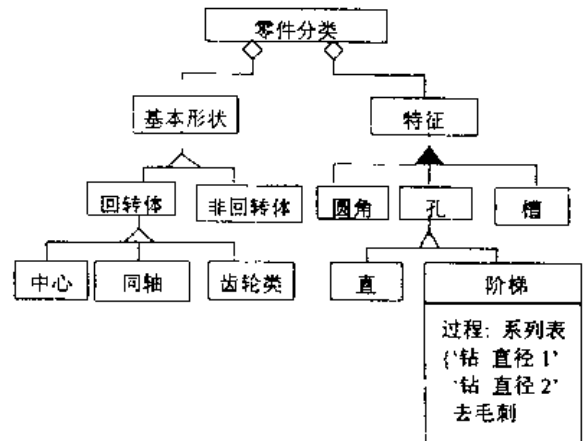


图 42.4-13 基于聚合语义抽象对零件的分类

GT 的 C&C 模型是一个强有力的工具，它们的使用应该能给企业带来可观的效益。但在实际应用中也有局限性 (有些甚至严重地影响到最终的应用)。首先是缺少柔性，如果要在码位中加入附加的信息则需要对整个编码定义施行“手术”，从而，导致时间和成本的浪费。这也是为什么在编码定义中常常留有空码位的原因。其次，是这样的编码定义不能满足产品活动的全过程的应用需求，或者说难以按照上面的模型构造满足产品活动全过程的应用需求，如果一定要做到，则码位势必很长，这又不便于使用。还有，上述应用比较广泛的三种 C&C 模型只提供了信息的某个范围的描述 (如  $0.5L < D < 1$ )，但不能提供原始值的准确描述 (如  $L/D = 0.75$ )，这会带来这样几方面的问题：第一，不能通过 C&C 模型中的订单定义检索到某个条目的准确值；第二，一旦发现所设计的 C&C 模型在某个方面不能满足要求，则必须对它进行重新构造；第三，属性范围的设定会因设计师的理解不同而不同。最后，已有 C&C 的研究集中于零件，对于面向企业不同层次的产品资源的分类定义尚未涉及，而这是定义与重用这些产品资源所不可缺少的。

伴随着信息理论和计算技术研究的深入，数据模型与 GT 的集成必将为 GT 的应用增添活力。数据模型是关于物理世界应用域所关心的部分属性的抽象，它可以描述数据对象的完整细节。面向对象模型 (OOM) 将数据对象的属性和操作封装在对象和对象类中，基于对象关于对象类定义的继承性可以表达不同层次的一类对象的继承关系，同时，基于对象的抽象性则可以定义任何类型的分类信息。本节基于 OOM 对象和对象类的特性，研究对象分类定义框架，为分类企业不同层次的产品资源提供了可行的解决方案。

### 3.2 对象分类框架

研究RPM的直接动机,是用企业成熟的产品资源来支持合理、快速的变型设计。产品重组后的规范化、标准化的产品资源,必须以有效的、合理的标识供产品活动全过程所共享,正因为这样,一个合理的产品资源分类框架应该具有这样的一些基本特征:

首先,以产品结构形状为最终的唯一标识。产品构成的结构形状作为它们的物理表达,其实现是与时间、成本、质量竞争三要素密切相关的。为了更好地向设计工程师提供规范化、标准化的企业产品资源的信息,一个以产品结构形状为最终标识的分类及编码(C&C)模型是必须的。也就是说,这样的C&C模型,不是从产品活动过程的下游部分的角度来构造产品资源的分类模型,而应该是从产品活动过程的源头——设计过程可重用产品资源的需求角度加以描述。因为只有这样,才可有效地提供给设计工程师足够的信息,使其快捷、有效地检索到企业产品资源。

其次,以对象类的属性定义构造分类空间。结构形状的相似,决定了对象功能抽象的相似(如齿轮轴类零件,其功能可抽象为以齿啮合的原理实现能量与运动的转换)。反过来,功能抽象的相似又为结构形状的相似定义了相关的映射约束(见本章第一节)。因此,有理由这样说,为了实现某个抽象的功能,其一类结构相似的对象应该具有很好的几何与非几何的相似性。基于此,将结构相似的对象中的那些结构相同的对象定义为一个集合(即将结构属性定义相同的对象定义为对象类),这无疑对建立与产品活动全过程的各子过程相对应的过程子模型有重要意义。

再者,对不同层次、不同类型的产品资源具有普遍的适应性。在RPM的理论体系中,强调关于零件对象、部件对象、产品对象的规范化、标准化的信息化的定义与表达,因此,支持RPM的C&C模型,应该覆盖关于这些所有对象的标识,而不仅仅是传统C&C模型中只是对于零件对象的标识。此外,这些对象作为企业资源,从标准化定义级别,从购买/自制的优先级别等诸方面反映了快速响应变型设计的需要。关于产品资源的A、B、C分析(A类:新设计零部件;B类:变型设计零部件;C类:标准与外购零部件)表明,A类产品资源无疑是属于那些对产品的交货期、成本、质量等有决定性影响的对象。因此,只有在C&C模型中反映了企业的产品资源属性的这种优先级,才是关于企业产品资源的合理的、完整的分类与定义。

还有,对不同活动过程相关的产品资源具有普遍适应性。为了加快订单规划的全过程,必须强调对企业产品资源的重用。因此,在RPM的研究中,除了设计

子模型外,其他几个子模型也必须要保证其产品定义的规范化、标准化。这就要求,RPM中的C&C模型除了表达设计子模型的分类定义外,还要能有效表达工艺子模型、制造设计子模型和成本设计子模型的分类定义。显然,这样的C&C模型对不同活动过程相关的产品资源必须具有普遍的适应性。

最后,层次的简化。传统的C&C模型面向的是族,几何相似是构造族框架的基本准则。但是,一个非常普遍存在的问题是:为了分类定义的完备性、唯一性,分类层次在深度方向难以统一定义,结果是编码的位数不得不加长。前已述及,数据模型与GT的C&C的集成,将有助于完备、唯一地表达分类定义。这样,由于数据模型的引入,一些传统的C&C模型中必须由数来标识的分类定义,如材料、工艺类别、精度等级等,就能够定义在数据模型中,这样,简化分类层次,减少在实用中由于层次的增加而带来问题解决的思路的复杂。

上述的讨论表明,基于GT分类原理,根据RPM中问题的侧重及求解的思路,应该能有一个合理的C&C模型,使得它能支持面向订单生产的快速响应的全过程。理论上,这样的分类使得一类结构相同的对象在订单生产的不同子过程中,都可以有相对应的对象子模型,从而有限地解决了与设计子过程同步的其他子过程变型的分类标识需要。在RPM的研究中,对象分类框架(OCF)的理论是关于企业产品资源分类定义。

对象分类框架(Object Classification Framework, OCF)是基于产品定义对象(PDO),根据企业一级的产品资源中对象类的功能属性、结构属性及形状属性的抽象,对产品资源进行分类标识的体系。

由上面的定义知道,OCF作为企业一级的产品资源的分类定义体系,它必须覆盖不同层次、不同类型的产品定义,必须唯一地、完备地定义这些产品资源的分类。PDO作为对产品构成本身的抽象,所定义的对象和对象类通过属性定义和操作定义表达了这种抽象。OCF基于对象和对象类的抽象特性,截取功能属性、结构属性和形状属性定义一个对象类的分类。逻辑上,OCF有四个层次,它们是:资源属性层(RAL)、组对象类层(GOCL)、族对象类层(FOCL)、对象类层(OCL)。OCF的语义描述如图42-4-14所示,

其中,资源属性层(Resource Attribute Level, RAL)从标准化级别、购买/自制状态、对象类的结构层次属性等关于企业的产品资源的分类定义。

对一个企业来说,在产品资源的类别中,覆盖了标准的构件,包括国际标准的构件(Int. Standard Components)、国家标准的构件(State Standard

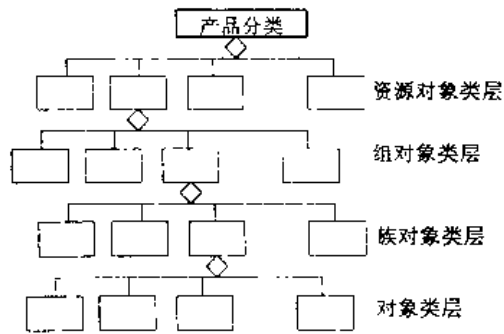


图 42.4-14 对象分类框架

Components)、行业标准的构件 (Trade Standard Components) 及企业标准的构件 (Enterprise Standard Components); 覆盖了非标准的外购构件 (Purchased Components); 覆盖了不同层次的产品构成, 包括零件、部件及产品; 覆盖了工装, 包括夹具、量具、卡具、刀具等等。这些不同类别的产品资源, 涉及订单规划过程中可重用资源的优先性的确定, 因此, 在 OCF 中将这样的资源属性的划分作为首要定义。这样的划分保证了相应的产品数据管理 (PDM) 策略可强化对企业产品资源的定义和重用的管理, 同时, 也为产品重组 (Product Reengineering, PR) 中优选对象类提供了有意义的指导。图 42.4-15 描述了资源属性层 (RAL) 的部分对象类。

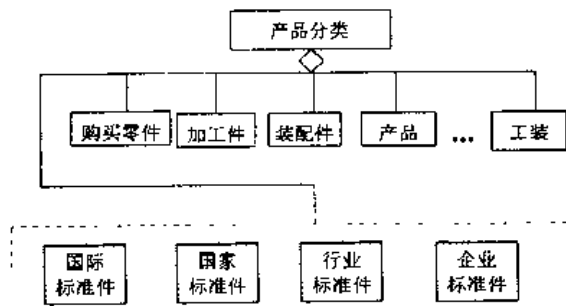


图 42.4-15 资源属性层

组对象类层 (Group Object Class Level, GOCL) 从抽象的功能定义、抽象的行为定义对 RAL 中一个对象类的对象分类定义。

在 RAL 中定义的资源属性的对象类, 从它的全部对象的抽象功能定义、抽象行为定义出发, 将具有相似的抽象功能定义、抽象行为定义的那些对象划分为一个对象类, 并被定义为一个组对象类 (Group Object Class, GOC)。GOCL 分类定义的 GOC, 其本质属性在于功能抽象、行为抽象一级上的相似性。例如, 企业自制零件 (Manufactured Parts) 按功能抽象、行为抽象可定义板状组对象类、齿轮组对象类、轴组对象类、箱体组对象类等等。在另一方面, GOCL 中的每个 GOC 继承了它的超类 (Superclass) RAL 的定义, 因

此, 它具有与它的超类相同的资源属性分类定义。GOCL 的分类示例如图 42.4-16 和图 42.4-18 所示。

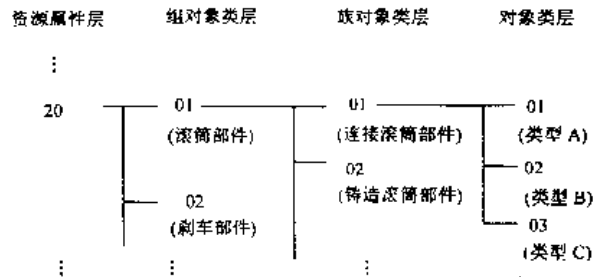


图 42.4-16 基于 OCF 的起重机滚筒部件分类

族对象类层 (Family Object Class Level, FOCL) 从结构属性的相似对 GOCL 中的一个对象类的对象分类定义。

在 GOCL 中, 每个 GOC 都将抽象的功能属性和行为属性作为自身的本质属性, 它的对象的结构属性的相似性是 FOCL 中分类定义的唯一准则。在 FOCL 中, 构成一个族对象类 (Family Object Class, FOC) 的全部对象, 一方面继承了它们的超类—GOC 的抽象的功能属性和行为属性, 另一方面它们的结构属性是相似的。例如, 作为“齿轮”这一 GOC, 它从结构相似性可以有齿轮轴族对象类、齿轮族对象类、锥齿轮族对象类等。以齿轮轴族对象类为例, 其结构相似性表现在所有齿轮轴的基本构成—结构单元的相似。FOCL 的分类示例如图 42.4-16、18 所示。

对象类层 (Object Class Level, OCL) 以结构属性的相同为准则对 FOCL 中的一个对象类的对象分类定义。

在 OCL 中, 每个对象类继承了它的超类 FOC 的结构相似的分类定义, 将相同的结构属性作为自身的分类定义, 这样, 它的所有对象都具有相同的结构属性。由于对象属性的继承性, 这些对象的抽象的功能属性和行为属性应该是相似的, 这就决定了它们拥有相同的抽象的形状属性和非形状属性。不难看出, 这样的分类定义的对象类所具有的结构相同的特性, 为 RPM 的逻辑相关的子模型共享对象的属性定义提供了重要的理论基础。同时它又为信息环境下定义与重用对象类的属性定义创建“变型”对象提供了重要的理论基础。在本章中, 对象类有时也称为实例对象类, 它们的对象有时也称为实例对象, 在此特作说明。OCL 的分类示例如图 42.4-16、18 所示。

上面的讨论表明, 本章提出的对象分类框架 (OCF) 所定义的 RAL、GOCL、FOCL 和 OCL, 通过四个不同层次的抽象由一般到具体对企业产品资源的分类作了定义。OCF 侧重于对企业产品资源的抽象定义, 从资源属性类别的抽象、功能和解的原理的抽象直

到结构的抽象，这不仅符合人们对客观世界问题的本质的认识与理解，而且，这样的抽象定义使得产品资源中不同层次、不同类型的对象类可以统一于一个框架中。这样带来的好处是很明显的，在重组企业的产品资源后，可以定义一个数据字典（Data Dictionary），将在企业的 OCF 中涉及的所有的分类层次的对象类作出规范化的命名。当工程师们使用熟悉的抽象命名去快速检索相关的产品资源时，最佳路径必然是选择该企业的 OCF。此外，这样的 OCF 在一个有效的产品数据管理（PDM）策略支持下，可以做出很多有用的事情，如基于对象类的类属模型（OC GM）进行的 CAD/CAPP 集成；又如基于 OCF 进行的快速检索；还有基于 OCF 的在 PDM 策略支持下管理产品资源的定义与重用等。有关内容在本章不作详细讨论。

OCF 作为 RPM 理论体系中的构成之一，为规范化、标准化的分类企业信息化产品资源提供了可行的解决方案。作为对企业产品资源的完整分类体系，OCF 不仅仅对产品的基本构成——零件类作了定义，而且对部件类、产品类也作了定义。这样，从 OCF 覆盖的范围来讲，它很好地解决了在 GT 的 C&C 模型中未予涉及的部件、产品的分类。因此，OCF 为企业提供了关于全部产品资源的完备的分类定义体系。同时，采用基于对象建模技术（OMT）的基本原理，对不同层次的对象类从资源属性、功能属性、行为属性及结构属性加以抽象，从而保证了 OCF 分类定义的唯一性。还有，OCF 是基于企业产品资源的对象类的属性定义抽象的结果，因此，对于不同的企业、不同的产品资源的类型的应用问题的求解，具有很好的适应性。这也就是说，不同的应用只是涉及领域内的专业知识不同，而并未涉及分类体系的重新构造，因此，它不存在因不得不

重构分类体系而造成的时间、财力、人力上的浪费。再者，OCF 的分类体系是与事物特性原理的定义相一致的，它遵循相关的国际标准、国家标准，因此，它的分类定义资源的可重用性将长期得到保证。最后，基于 OCF 的对象类类属模型的理论，又为订单生产活动过程内统一地定义产品数据，并且有效地共享这些数据提供了重要的基础。可以说，如果将 RPM 作为快速响应生产过程的理论基础，则 OCF 是这个过程取得成功的组织保证。

图 42.4-1 中根据产品的构成关系，将产品的信息资源分为产品类、部件类和零件类三大对象。对象分类框架（OCF）则根据对象类的功能属性、结构属性和形状属性的抽象，进一步对产品资源进行分类标识。考虑到企业广泛使用标准件和外购件，并且拥有大量的工装和设备，所以采用一个如图 42.4-17 所示的四层分类方案是适当的。图中，资源属性层定义对象类的标准化级别、购置/自制状态和结构层次属性的类型，可分为 00—标准件/外购件；10—零件；20—部件；30—产品；…等类型。在资源属性层的同一个类型中，组对象类层定义功能属性和行为属性（概念模型）相同的一组对象类，如螺栓和螺钉、垫圈和环、齿轮、轴和箱体组对象类等。在同一个组对象类中，族对象类层则定义结构属性相似的一些对象类，如在齿轮组对象类中，可以分别定义齿轮轴、直齿轮、斜齿轮、锥齿轮等族对象类。而对于同一族对象类，又可按结构属性相同的原则，再定义下一层的实例对象类，在每一个实例对象类中，对象之间形状相同，其差异只在于形状值的大小不同。这种对象分类框架，从企业资源属性类别的抽象、功能和行为（概念）的抽象直到结构的抽象，通过四个层次对企业产品信息资源由一般到具体地进行分类定义，使

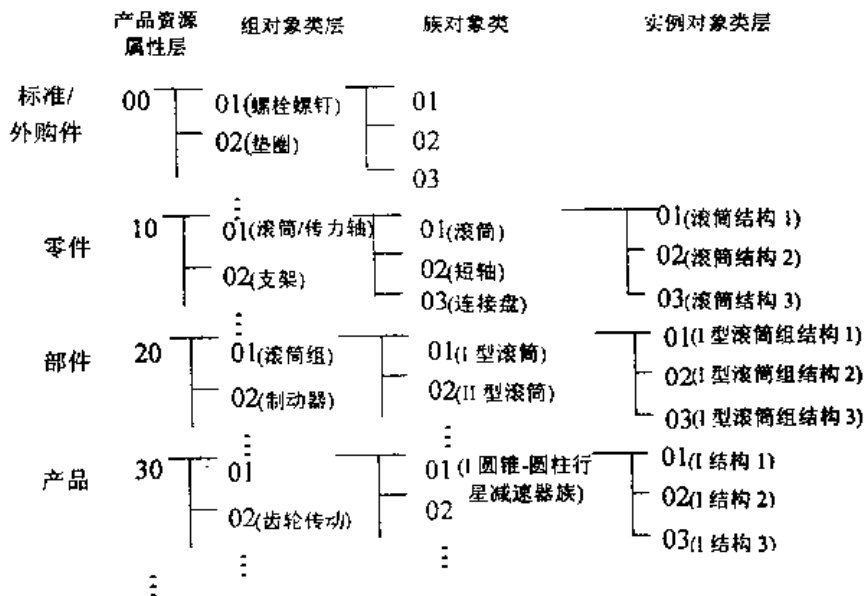


图 42.4-17 产品信息分类框架



产品资源中不同层次、不同类型的对象类可以统一于同一个框架之中,从而为准确、快捷地分类、管理产品信息资源提供技术保证,也为在变型设计中有效地重用企业产品信息资源提供合理的方案。

#### 4 基于关系的族类属模型

在本章的第二节,研究了产品定义对象(PDO)的理论,对产品资源中不同层次、不同类型的产品构成本身的数据和数据关系加以抽象,用对象和对象类定义与表达这种抽象。在上节的讨论中,研究基于PDO的对象分类框架(OCF)的理论,完备地、唯一地分类企业的不同层次、不同类型的产品资源。引入对象和对象类,实现了企业产品资源中不同层次、不同类型的对象和数据关系的高度抽象。自然,在先进的信息环境下,定义与表达这样的抽象应该也是关系型产品模型(RPM)理论体系的构成。在这一节中,将基于PDO和OCF的基本理论,基于对象类的抽象特性,研究对象类的类属模型(OC-GM)的理论以定义与表达这样的抽象。OC-GM以对象类作为语义单元,覆盖零件对象类(PartOC)、部件对象类(AssemblyOC)和产品对象类(ProductOC),研究它们的属性定义和操作定义在信息环境下的定义与表达的理论和方法。

##### 4.1 对象类的基本特性

由本章第一节知,产品重组后的产品数据和数据关系是企业产品资源的规范化、标准化的定义。如何将这种定义在信息环境中有效地表达,并能作为信息化的可重用产品资源在订单产品的快速响应变型设计中使用,需要涉及这样两方面的问题:首先,在本章第二节的讨论中探讨了数据与数据关系的定义与表达,在PDO的框架下的元对象模型是关于产品构成的数据和数据关系的描述。只有将数据关系以元对象模型“装配”形式与数据集成,才能在不同的参数作用时保证可靠的变型(及保证重用状态的定义)。其次,作为信息化的可重用资源,关于定义状态和重用状态的正确定义无疑是必要的。然而,合理的管理也是必不可少的,在上一节中对象类分类框架(OCF)从抽象到物理的演绎对企业产品资源进行分类,这样的分类体系注重于订单生产上游的设计过程,并且在合理的PDM策略支持下使其能覆盖全过程。详细考察一下OCF中对象类层(OCL)的对象类的基本特性,对于理解RPM中的零件模型、部件模型及产品模型的构造及实现是有意义的。

由上节的定义可知,对象类层(OCL)是关于族对象类层(FOCL)的一个族对象类中结构属性相同的对象的分类。基于OOM的基本原理,OCL中每个对象

类是FOCL的一个族对象类的子对象,因此,它完全继承了FOCL中定义的全部属性,同时,也有自己的属性定义。现以图42.4-18所示的齿轮轴对象类为例,说明对象类的属性继承关系。在图42.4-18中,齿轮作为OCF下一个组对象类,它的子对象定义在FOCL,包括齿轮轴、齿轮、锥齿轮、蜗轮等族对象类(Family Object Class, FOC)。对于这里的每个FOC,它继承的是齿轮组对象类所定义的功能属性(如“实现广义运动的转换”,“实现广义力的转换等”)、行为属性(如采用齿啮合实现所需功能)。除此之外,它又定义有自身的结构属性,如齿轮轴FOC,它的结构属性中可以定义有“采用齿轮与传动轴同体结构”,“采用左、右轴颈支承结构”等等。这里的每个族对象,都有在OCL的一个或多个对象类作为它的子对象。仍以齿轮轴FOC为例,它的子对象可以是左/右无输入/输出轴段、左端有输入轴段、右端有输入轴段等等。这里,每个对象类作为齿轮轴FOC的子对象,继承了它的全部属性,同时又定义有自己的结构属性。另一方面,每个对象类又同时定义有一个或若干个对象。它们继承了对象类的完全相同的属性定义(几何属性与非几何属性),属性的值决定了它们之间的差异。图42.4-18是齿轮轴对象类所继承的不同层次的属性定义的语义描述。

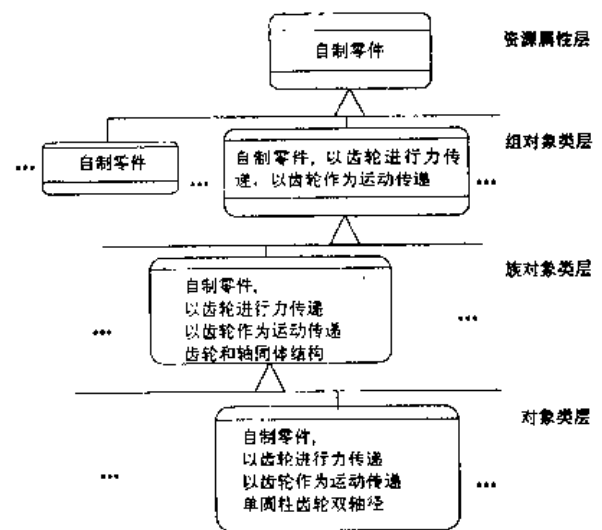


图 42.4-18 齿轮轴对象类属性继承关系

OCF中的一个对象类,是由多个具有相同属性定义的对象集合。在这些属性定义中,既有从不同层次的超类中直接(或间接)继承的,又有自身的属性定义。在OCF的框架下,从不同层次的超类中直接(或间接)继承的属性定义为语义属性(即隐式信息),如功能属性、行为属性。在自身定义的属性中,则是关于对象类自身与构成有关的定义,它们可以是结构属性、精度属性、材料属性(可选)、质量属性、技术属性及管

理属性(详见下面的讨论)。另一方面,作为产品定义对象(PDO)关于产品构成本身的抽象,对象类的对象的可重用是关于它们的属性定义进行操作的结果,这样,与重用状态有关的信息(即数据关系)应该作为对象类抽象定义的完整组成。在RPM的理论体系中,对象类的类属模型(OC GM)将对象类作为不同层次产品资源的基本语义单元,研究关于零件对象层次、部件对象层次以及产品对象层次的作为可重用资源的数据和数据关系的定义与表达。

## 4.2 对象类的类属模型

对象类的类属模型(Object Class Generic Model, OC GM)是在关系型产品模型(RPM)的理论体系中,基于产品定义对象(PDO)和对象分类框架(OCF),定义和表达对象类的属性定义和操作定义的语义模型。

在PDO和OCF的定义框架下,对象类作为不同层次、不同类型的产品资源的抽象,它的属性定义和操作定义是关于这些产品资源的数据和数据关系抽象表达。在产品数据的定义中,数据作为设计活动的结果,是由一些显式表达的信息构成的;数据关系作为设计活动的过程和取值的关系信息,是隐式地定义在数据信息中的。我们知道,RPM是以研究规范化、标准化的产品资源的可重用为中心,因此,必须将PDO和OCF的定义框架下的一类产品资源的数据和数据关系加以抽象,并表达为对象类的属性定义和操作定义。由于对象和对象类的这种高度抽象性,本文才得以基于一个语义模型-OC GM来统一地定义与表达不同层次的产品资源的数据和数据关系。

基于PDO和OCF的定义框架,一类产品资源在定义状态的定义相对比较简单,包括:

- 语义功能定义;
- 结构构成的抽象定义;
- 形状描述的抽象定义;
- 非形状描述的抽象定义;
- 技术要求的抽象定义;
- 视图表达的抽象定义;
- 管理信息的抽象定义。

在重用状态下,它们的定义更多地包括了与构造期望的形状有关的数据和数据关系,它们是:

- 在空间任意参考坐标系下参与更高一层的装配必须的装配基准的抽象定义;
- 在空间任意参考坐标系下保持形状定义的相对拓扑关系的抽象定义;
- 在空间任意参考坐标系下取得期望形状定义的取值关系的抽象定义;
- 在空间任意参考坐标系下取得期望的装配位置

的确定装配基准的操作定义;

- 在空间任意参考坐标系下在装配基准取得期望的空间相对位置的操作定义;

- 在空间任意参考坐标系下取得期望的形狀的操作定义。

基于对象和对象类的抽象特性,对上述一类产品资源的定义抽象,并表达为对象类的属性定义和操作定义。在定义状态下,它们是:

- 功能属性;
- 结构属性;
- 形状属性;
- 参数属性;
- 技术属性;
- 视图属性;
- 管理属性。

在重用状态下,它们是:

- 基准属性;
- 旋转属性;
- 隐藏属性;
- 索引属性;
- 定准操作;
- 旋转操作;
- 隐藏操作;
- 索引操作。

关于这些属性定义和操作定义的对产品对象类、部件对象类和零件对象类的详细定义和讨论,在此不展开。

对于这些属性定义和操作定义进一步讨论,会发现它们可以归纳为下面的五类信息:

- 与OCF分类定义有关的分类信息,表达了对象类的功能属性和结构属性;
- 与形状表达有关的形状信息,表达了对象类的结构属性、形状属性、基准属性、旋转属性、索引属性、定准操作、旋转操作、索引操作;
- 与非形状表达有关的参数信息,表达了对象类的参数属性、技术属性、索引属性、索引操作;
- 与形状和非形状的定义有关的视图信息,表达了对象类的视图属性、管理属性、隐藏属性、隐藏操作;
- 与对象类的定义和重用有关的管理信息,表达了对象类的管理属性。

对于第一类信息,它的直接表达体现在分类定义中。因此,可用分类码来定义,其解释可用相应的PDM策略实现。对于第五类信息,管理属性必须与PDM策略的相应过程匹配。这两类信息的定义与管理直接与PDM的策略有关。对于第二类信息,集中描述的是对

象类的形状定义,因此,在关系型CAD平台中构造形状模型。重要的是,这样的形状定义,不仅是定义状态,而且也是重用状态。所以,元对象模型(MOM)将在这样的定义过程中扮演非常重要的角色。对于第三类信息,它们是对象类在重用状态下形状变型必需值的定义,同时又是与取值的定义过程有关的。事物特性是在一个对象组中区别和标识一个对象的决定性的特性,因此,可将它们按基于事物特性的原理加以定义。

视图是形状定义和非形状定义的表达载体,一定的视图是对象类完整定义所不可缺少的。构造视图定义的信息模型是第四类信息表达必需的。综上所述,OCGM的逻辑构成应该包括:

- 定义与表达一个对象类的分类信息的分类定义模型(CDM);
- 定义与表达一个对象类形状信息的形状定义模型(FDM);

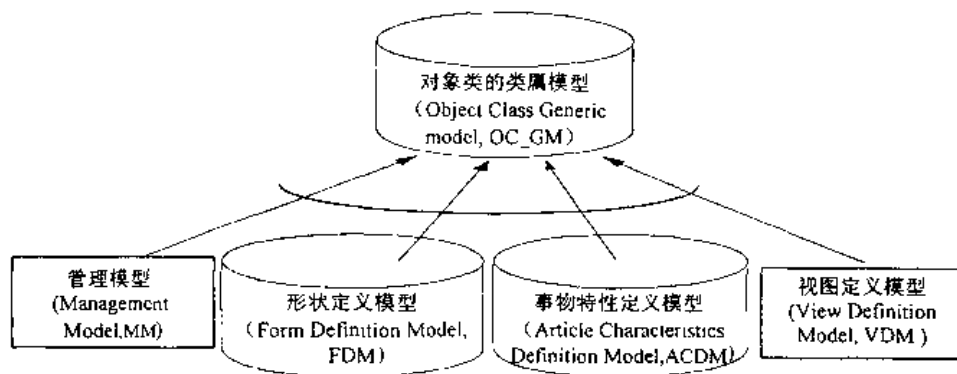


图 42.4-19 对象类的类属模型

●定义与表达一个对象类非形状信息的事物特性定义模型(ACDM);

●定义与表达一个对象类视图信息的视图定义模型(VDM);

●定义与表达一个对象类管理信息的管理模型(MM)。

OCGM的语义表达如图42.4-19所示。

本节研究的对象类的类属模型(OCGM),是基于对象和对象类的抽象性,定义和表达产品资源中不

同层次的对象类的属性定义和操作定义的语义模型,它的逻辑构成包括分类定义模型(CDM)、形状定义模型(FDM)、事物特性定义模型(ACDM)、视图定义模型(VDM)和管理模型(MM),它们从不同侧面,借助不同的信息平台,实现了统一地定义和表达不同层次的对象类的属性定义和操作定义。如果说,在RPM的理论体系中,PDO和OCF是基于抽象特性来讨论不同层次产品资源的数据和数据关系,那么,OCGM则是实现这种抽象的理论和方法。

## 第 5 章 基于实例推理的快速响应变型设计

近年来在国际上出现了将基于实例设计 (Case-based Design, CBD) 系统的研究和开发引入设计领域的趋势。CBD 系统是新兴的人工智能技术 CBR (Case-based Reasoning) 基于实例推理技术在 CAD 领域应用的结晶。CBD 即是组合或修改已存在的设计结果, 产生满足新设计要求的变型设计的过程。CBD 是在 1988 年 Kolodner 主持的第一届国际 CBR 技术研讨会上正式提出的。从 Navinchandra 的 CYCLOPS——第一个真正的 CBD 系统开始, CBD 技术的发展也只有十余年, 但近几年国外作了大量的研究工作, 并开发了许多原型和商用系统, 这不仅大大地推动了智能 CAD 技术的发展, 而且充分证明了 CBD 在智能 CAD 的研究和实践中的成功性。国内在 CBD 方面的研究尚处于起步阶段, 因此开展 CBD 的理论和应用研究不仅意义重大, 而且十分迫切。

### 1 CBR 概述

CBR 是人工智能领域新崛起的一种重要推理技术, 是近十几年发展起来的区别于基于规则推理的一种推理模式。早期的专家系统多是应用基于规则推理 RBR (Rule-Based Reasoning, RBR) 和基于模型推理 MBR (Model-Based Reasoning - MBR) 的技术方法, CBR 方法并未受到重视。直到后来专家系统出现了知识获取“瓶颈”和“知识爆炸”等问题, 才使得一部分研究者重审 CBR 的重要性。CBR 思想基于认知心理学原理, 是一种相似推理方法, 其核心是提供一套方法帮助人们应用过去的经验和实例解决当前的相似问题。

人们常采用基于实例的推理来提高解决问题的效率。例如, 当一个医生面对着一个具有多种症状的病人时, 如果医生曾见过具有类似症状的病人或阅读过有关病例报告, 他(她)往往就回忆起旧的病例并参考旧病例的诊断, 以作出对新病人的诊断。如果单纯依靠医学知识以及各种测试手段来提出一个类似的诊断, 就需要花费很长时间, 而利用过去的经验可以使医生节省大量的诊断时间。当然这个旧的诊断不一定完全适合于新的情况, 医生还需要对它加以证实和修改, 但不管怎样, 回忆旧的实例能使医生更快地得出准确的结果。事实上, 基于实例的推理在人类的各种活动中普遍地存在着, 人们如此擅长并如此普遍地使用基于实例的推理, 以至于它已成为人工智能研究领域的重要课

题之一。

在 CBR 中, 实例库模拟了人脑的记忆, 其中存储了一些过去的有关经历(实例)。这些实例按一定方式组织, 以便在需要的时候能被迅速取出。回忆过程对应了 CBR 中从实例库中检索出相关实例的过程; 被检索出的相关实例可能与新的情形不完全一致, 这时需要对旧实例的某些特征进行修改, 使它适合新的情况, 以得到对新情况的预测或新问题的解。对新情况的预测或新问题的解不一定完全适合实际情况, 它们还需要得到检验, 如果检验发现它们与实际不符, 则需要对它们加以修正; 最后新的实例(包括正确的解答)被存入实例库中, 同时新实例的索引也被建立和存储, 这时系统学到了新的知识。这个过程就是基于实例的推理和学习过程, 图 42.5-1 表示了这个过程。下面具体说明各个过程:

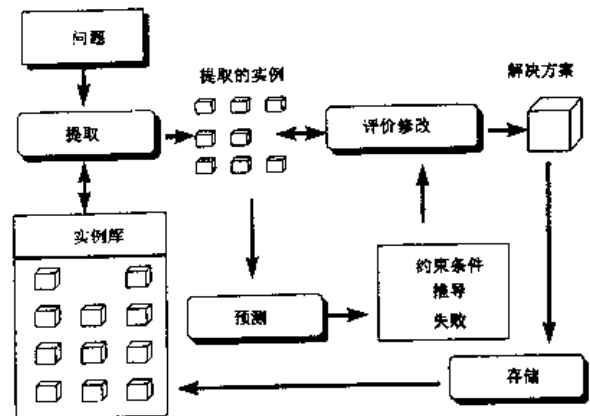


图 42.5-1 实例推理的基本过程

●提出问题: 输入待解决的问题的要求、初始条件以及其他相关的信息。

●检索和提取实例: 根据问题要求、初始条件, 从实例库中提取一个或一组与当前问题相似的实例。

●评价修改实例: 从相似实例中找出最相似的实例或多个实例的组合, 形成目标问题的可能解决方案, 并通过目标方案的评价修改来满足当前的设计要求;

●存储实例: 问题解决之后, 当前的解即可作为新的实例存入实例库, 实现自学习的能力。

CBR 推理技术引入设计领域是为了克服传统的基于规则的专家系统和基于模型的推理系统的缺陷。知识系统在设计领域已应用多年, 特别是专家系统 (Expert System, ES) 在 CAD 中的应用确实推动了智

能CAD的发展,但这些应用系统均不可避免地存在以下问题:

- 知识获取问题 有效的知识获取仍是提高ES推理能力和实用性的瓶颈问题,人类的许多知识、有价值的设计经验是很难准确表达的,虽然研究者们试图建造一些知识获取工具来帮助解决这一问题,但成效不明显;

- 知识表示问题 ES无法用目前常用的知识表示方法(如一阶谓词、语义网、框架和规则等)准确地描述人类的经验性知识;

- 进化能力差,即自学习的能力差;

- 实施专家系统也是一个很复杂的过程,需要专门的技术,并往往需要许多人的工作;

- 系统运行过程难以维护。

而CBR推理技术则可避免这些问题,并具有如下优点:

- CBR推理不需要精确的领域模型,因而知识的提取变成了一个收集实例的过程;

- CBR推理的执行过程简化为确定描述实例的典型特征,这比产生一个精确的模型或众多的规则要容易的多;

- 应用数据库技术,可以解决大量信息的管理问题;

- CBR系统能通过把新的知识转换为实例而实现自学习,这样使系统的维护容易。

### CBR的发展历史

20世纪70年代中期,就已出现一些体现CBR思想的计算机程序。随着关于记忆的认知理论不断成熟,CBR的计算模型不断完善,并被迅速应用于建立智能系统的技术中。到了80年代,则出现了形形色色的基于实例的系统,CBR已被广泛应用于各个领域。

下面是一些最早的能体现CBR思想的计算机程序:

- 1) HACKER程序(Sussman1975),这个程序体现了CBR的许多原理,HACKER的解答库和调试过程分别类似于实例库和规划修复(planrepair)过程。HACKER的认知前提是经验学习,这正是CBR的核心思想,HACKER中的情景(实例)是计算机程序,而不是更广泛的人的经历。

- 2) SAM(Cullingford 1978, script应用者机制)和FRUMP(Dejong 1979,快速阅读理解记忆机制),程序中使用静态知识结构script,但没有说明如何获取知识,而且程序不能记忆。

- 3) CYRUS和IPP(Kolodnet 1980, Schallk & Kolodner 1979, etc.)讨论了情景知识的记忆组织。

CYRUS模拟对外交官Cyruws的自传性事件的记忆,IPP则通过阅读关于恐怖活动的事件,形成概括性知识,并用来理解新的故事。

20世纪70年代后期对情景记忆模型加以模拟的研究主要是在自然语言理解领域。而在这个时期,另一人工智能研究课题已发展为主要应用领域,即基于规则的专家系统。这些专家系统的基本知识是规则,一个典型的诊断任务就需要几百条规则,随着对专家系统的经验增加,基于规则的系统(Ruled-based System)的缺点越来越明显:

- 知识获取问题。前面提过,建立专家系统时规则难以获取,专家们常常难以列出他们所用的规则。知识获取问题成为建立专家系统的瓶颈问题。

- 基于规则的系统没有记忆,它们不能记住曾经解决过的问题,并且一次又一次地重复已做过的事,其效率是很低的;另外这些系统也不会记住它们曾经犯过的错误,这导致它们必然重犯错误,其正确性也是低的。

- 基于规则的系统鲁棒性不好,如果向系统提出的问题不能匹配任何规则,那么系统就不能解答。

将基于规则的系统同人类专家相比,可以发现:

- 1) 专家的特点是具有经验,专家从经验中获取知识,通过吸收新的实例来增加知识。

- 2) 专家能记住他们成功或失败的经历,利用成功的经历引导将来的实践,同时避免失败的经历再次发生;

- 3) 人类专家可以进行类比推理,专家可以回忆在这方面或那方面与目前情形类似的实例,得出一个综合的解方案。

以上比较表明,基于实例的推理在以上三方面与人类专家更接近,克服了基于规则推理的三个问题,因而基于实例的系统成为相对于基于规则系统的另一种可选方案。20世纪80年代以来,各种基于实例的系统(Case-based System- CBS)得到了迅速发展。

前面已提到过,CYRUS与IPP是CBS的原型,它们详细讨论了如何在记忆中组织实例,以及如何在实例库中检索实例,并利用检索出的实例理解新的情况。这些问题是CBS建造的核心问题。MEDIATOR、JUDGE及CHEF系统将CBR进一步用于问题求解、评估,以及规划与学习领域。其中CHEF系统已全面地讨论了实例的表示、检索、修改、测试、修正及从失败中学习等问题,是一个在技术上十分完整的CBS。这表明CBS的基本框架已经形成,建造CBS的技术已经成熟。

从20世纪80年代后期到90年代,出现了大量的CBS,这表明CBR已广泛应用于医疗诊断、法律、电路或机械设计、故障诊断、软件工程等各个领域。

从应用上看, 这些系统有这样一些特点:

#### (1) 解决的问题越来越复杂

CBS 逐渐应用于越来越复杂的任务。例如在设计领域, FIRST 系统只应用于简单的大梁设计, KRITIK 只在简单的电路设计和热系统设计上进行实验, 而 OPAMP 则发展到对复杂的电子电路进行设计。

#### (2) 应用领域的智能程度越来越高

例如 POLYA 将 CBR 技术用到几何定理证明领域, 模拟专家证明几何题的认知过程(而不是采用逻辑或代数方法), 这时计算机不再是只进行搜索和计算, 而是像人一样从图形开始着手证明。POLYA 将一些几何规划作为实例, 这些实例反映了专家的知识。再如 Project Visualization 系统, 以动画片中武士搏斗为试验场, 通过对过去这一过程或行为的观察来判断智能代理(agent)在某一情况下应采取什么样的行动。这个系统中的具体表示和投射与人的感觉十分类似, 类似于人的空间视觉想象。这两个系统都模拟了形象思维, 在很高的层次上模拟了智能。

#### (3) 系统规模朝大的方向发展

要使系统真正实用, 系统必然朝着大规模方向发展, 因为知识越多, 解决问题的能力就越强。一个例子是最近发展起来的 Corporate-Wide CBSs, 将 CBR 技术用于公司活动管理。有些系统将实例与数据库结合起来, 用数据库技术来管理实例库, 可见实例库的规模已十分巨大。

在各个领域中建造智能的 CBS, 促进了 CBR 技术的不断发展, 使 CBR 技术出现了一些新的特点:

#### (1) 实例的表示、检索及修改技术越来越复杂

早期的 CBR 计算机模型与 CBS 讨论了关于实例的表示、检索及修改的基本技术。随着 CBR 在更多更复杂的领域中的应用, 实例的表示、检索及修改技术越来越复杂。例如, 对于复杂的任务, 一个实例又可以分为许多小的实例组, 因而出现了单实例组(individual cases), 微实例组(microcases), 以及多实例组(multicases)三种形式。索引机制也由单一的索引机制发展到多种索引方式结合于一个系统中(如 BankXX 系统), 而且检索和修改机制还可以得到动态调整(如 CABOT 系统)。

#### (2) 结合领域知识

解决问题只凭经验是不够的, 还需要一些原理性知识或其它领域知识。早期的 CBS 已经结合了一些领域知识, 它们体现在索引规则、修改规则及修正规则上。近期的 CBS 还建立一些领域模型, 使它们参与推理, 从而更深地结合领域知识。如 KRITIK 系统使用物理元件的定性模型, 将它们应用于实例的检索、修改以及对设计的仿真测试中, 再如 Simoudis 和 Lee 提出

的生效模式(validation mode), 可在故障诊断中用来减少测试次数, 提高检索效率;

#### (3) 集成多种机制

CBR 是与 RBR (Rule-Based Reasoning) 不同的推理模式, CBS 克服了 RBS 的许多缺点, 近年来人们发现, 在某些领域将 CBR 与 RBR 这两种机制相结合更为有效, 典型的应用是在法律领域, 如 CABARET、GREBE 等系统, CABARET 系统说明在应用法律规则时, 使用 RBR 可以帮助理解意义不明确的术语, GREBE 系统则说明以前的实例可以帮助匹配规则的前提与案情, 而规则可以在实例中推出事实, 以增加实例间的相似性。

另一种集成机制是 CBR 与 MBR (基于模型的推理) 的结合。Goel 等人的 KRITIK 系统, 将 CBR 与 MBR 集成起来解决设计问题。设计对象的结构一状态-功能模型被用到实例库的检索、修改及仿真测试中。Rajamoney 等人提出的基于原型的推理(Prototype-Based Reasoning, PBR), 也集成了 CBR 与 MBR。它指出, MBR 适用于解决新问题, CBR 适用于解决经常遇到的问题, 但解决大的新问题时, 两种方法都遇到困难, MBR 变得难以控制, CBR 找不到相关的已有实例, 这时如果集成 CBR 与 MBR, 则可以有效地解决这类问题。PBR 将大的新问题看作一些熟悉的子问题的新的结合, 先用 CBR 发现和解决这些子问题, 通过结合这些子问题的解形成一个新问题, 再用 MBR 解决这个新问题。

#### (4) 强调系统结构

复杂的系统通常采用多种知识源及多种推理机制, 为使多种知识源及多种推理机制相互协调, 必须重视系统的结构。目前的一些系统表明, 对于多种知识源及多种推理机制的情况, 黑板是一种有效的结构, FIRST 系统和 FRANK 系统讨论了这种结构在 CBS 中的应用。

#### (5) 人机协作

以前的 CBS 大都着眼于“自主型”的计算机系统, 其应用几乎永远会令人失望, 不能达到人所期望或人能达到的程度, 一个新的思想是研制人机协作的 CBS。CBS 创始人之一 Kolodner 则互补地结合人和计算机的优势指出, 人擅长于从例子中形成预测及修改例子, 但人是健忘的, 记不住很多的实例, 而机器不会健忘, 可以记住大量的例子, 按一定索引组织后可以随时取出例子。因此可以建立一个机帮人的系统, 机器帮助人回忆出最恰当的实例, 而由人来作修改与决策。教学系统 CreANIMATE 及 GIT 的 Goel 等人形成的系统 AskJef 都体现了这个思想。人机协作的系统减少了建立自主型系统的难度, 可以发展为真正实用的系统。

CBR 推理方法和人在进行产品设计时的思维过程十分相似，因此使 CBR 技术在智能 CAD 领域中的应用迅速发展起来，使基于实例的设计 (Case-based design, CBD) 系统的开发和应用成为智能 CAD 的一个重要研究方向。

## 2 CBR 的基本问题

在基于实例推理的设计中，有两个关键问题：一是明确检索出和当前设计问题相关的设计经历和经验，二是确定这些设计中有哪些可以保留，哪些必须改变。

第一个关键问题包括对新问题设定索引、提取实例和选择实例。设定索引是提供特征属性集合以反映当前问题；提取实例是利用设定的索引对实例库进行搜索，提取出和当前问题具有相似索引的设计实例集合；选择实例是从这个相似实例集合中提取最相似设计实例。

第二个关键问题则包括实例的评价和修改，对提取出的最相似实例进行评价，确定不满足要求的部分，并对其实施修改，若修改后的实例满足新问题的设计要求，则提交设计并将这个实例存入实例库，完成最终的设计工作。图 42.5-1 表示了这个过程。

下面我们对这两个关键问题展开讨论，概要分析包含在其中的 CBR 的三个基本问题：实例库的建立、实例的检索和提取及实例的修改。

### 2.1 实例库的建立

实例库的建立实际上包含了实例的表示方法、索引及实例的组织结构等问题。实例库的建立和专家系统的知识获取相类似，同样具有一定的难度，且是一个病态过程。对 CBR 来说，实例库就是知识库。

实例库的建立有这样三个方面的问题：

- 每个实例的内容；
- 实例内容的表示方法；
- 实例库中实例的组织方法。

每个实例的内容是对已解决问题的基本描述，问题的关键是如何提炼出必要的信息来完整地表达已有问题解及其解的过程。我们认为对于存储于实例库中的每个问题解决方案，必须具有定义完备的需求说明、解方案的属性特征，以及如何获得这个解的过程的描述，而且当一个问题需求说明并不是十分完备时（这对产品设计问题是相当普遍的情况，问题的描述在设计开始之初并不是十分完备的，而是随着设计的进行而不断完善的），实例库必须提供相关知识以利于寻找和新问题相似的实例，这一般采取特征索引和优先权方法实现。

实例表示定义了实例内容的存在形式，可以采取特征属性一值对、整体-部分关系，或是特征属性网络等。图 42.5-2 表示了实例的三种存在形式。特征属性一值对表示方法最简单，但它反映的信息量也较小，

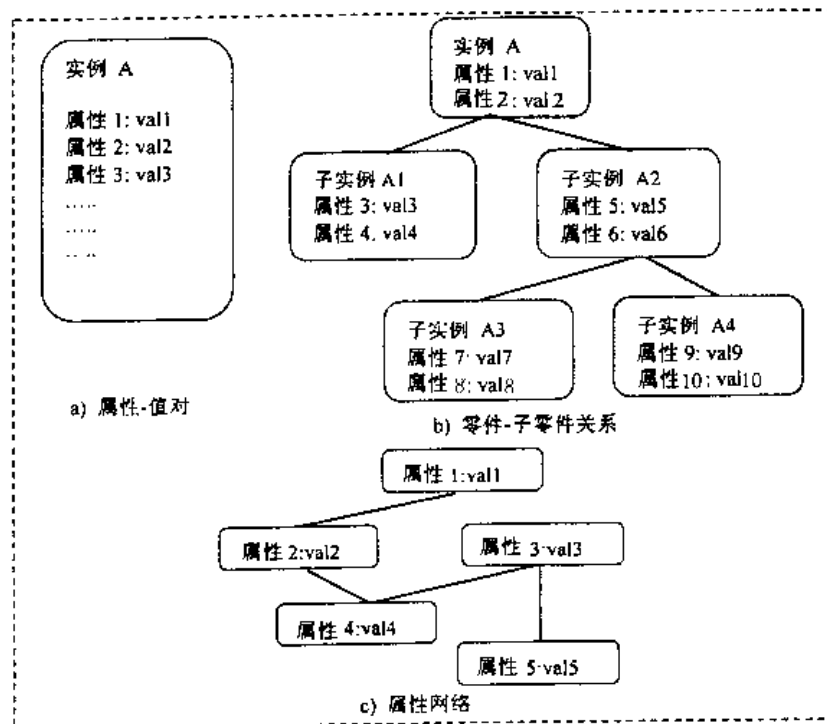


图 42.5-2 实例库中实例的表示

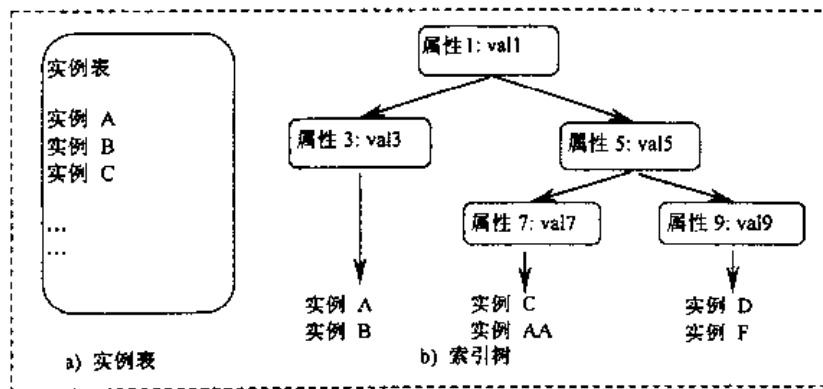


图 42.5-3 两种实例存储方法

整体一部分关系实例表示法及属性网络表示法，一般用于表示大型或复杂的实例，它们中均定义了一些关系信息。

实例库中实例的组织方法提供了一种机制，以在实例库中定位一个实例或其部分，它以合适的结构存储和组织实例为 CBR 系统快速、准确搜索相似实例提供保障。当实例库十分庞大时，良好的库结构尤其重要。组织实例最简单的方法是在链表中存储每个实例的名称和指向实例内容的指针，而当实例库较大时，采用“索引树”（树上的每个节点存储属于一个实例或多个实例的特征属性—值对）则可以减少需搜索的实例库空间。图 42.5-3 简单阐明了实例存储方法。

## 2.2 实例的检索和提取方法

从实例库中检索提取一个实例是根据新问题的描述进行模式匹配的过程，新问题可能被定义为一组特征属性值对，或一约束，或属性网等，在给定这些初始条件后，实例检索的过程是：建立索引——实例检索提取——最相似实例选择。

建立索引的过程是对新问题初始条件标准化的过程。从初始条件中提取能反映问题本质的特征属性作为索引，或对索引进行加工整理，如加权处理等，经过这个过程，新问题变为实例系统可以认可的模式。这样就可以输入实例系统，进行实例检索。

实例检索是相似实例的搜索过程，这是实例系统的关键，如何以最快的速度搜索出最相似的实例，它涉及相似实例的条件、搜索算法及终止条件等问题。

实例选择是从实例检索过程获得的一组相似实例中提取最相似实例的过程，通过各种判别标准的应用，可以将获得的相似实例进行排队，从而获取最相似实例。

## 2.3 实例的修改

实例修改是指通过对已有实例的修改或重组，生

成满足新要求的实例。实例修改涉及两个方面的问题：一是如何判别相似实例中不满足当前要求的部分，二是修改的方法。实例的修改要增加新的知识，在设计实例的修改中尤其需要如此。设计实例的修改必然也涉及知识推理问题，基于知识的实例修改在基于 CBR 的系统中是必然的选择。

实例修改是 CBR 系统的关键技术之一，也是最困难的一环。因而现有的许多 CBR 系统尽可能避免这个问题，如 Archie 系统仅向用户提供实例的存储和检索等功能，而由用户手工完成对设计实例的修改。

为了支持实例修改，大多数系统在 CBR 技术之外引入了其他人工智能技术，如约束满足技术、定性推理技术、基于规则的推理和基于模型的推理等，从而构成了混合型 CBR 系统。

许多 CBR 系统采用一组相互间有约束关系的设计特征属性变量来描述实例。对已有实例的修改便是从实例描述中找出不满足新设计要求的变量，修改之，并使修改后的变量赋值满足给定的约束关系。于是，实例的修改过程便是一个约束满足过程。典型的例子有 JULIA 和 CADSYN 等系统。CADRE 系统着眼于连续值变量，系统先将变量分成独立变量和非独立变量，约束冲突的修改仅作用于独立变量，非独立变量是随相关的独立变量的更新而更新的。系统采用尺寸归约方法支持对设计实例几何图形的修改。

PANDA、ASPII 和 BENTON 等系统采用 RBR 支持实例修改。实例修改的方法包括对单个具体设计实例的修改，对一组子实例的重组以及对抽象的模型的重新实例化等。

实现实例修改的另一种方法是结合基于模型推理 (MBR) 和 CBR。检索生成的实例作为问题的初始解，模型提供领域知识，支持对已有实例的评估，支持对其中不满足当前要求的部分的修改。在 KRITIK 系统中，实例修改被视为问题诊断和修正，实例评估被视为一个模拟仿真过程。



### 3 应用于快速响应变型设计的 CBR 关键技术研究

在变型设计中应用 CBR 技术,其关键问题仍是人工智能领域中关于 CBR 技术所研究的基本问题,我们将其细化为:

- 1) 实例的表示和组织;
- 2) 实例的检索;
- 3) 实例的修改;
- 4) 实例的存储。

实例的表示是基础,实例的检索通过搜索实例库,判别实例间的相似性,提取与需求实例最相似的实例,实例的修改是对提取的相似实例中不满足新设计要求部分的完善,而实例的存储则实现系统的自学习功能。

但设计实例具有本身的复杂性,需要表达的信息量大,而且信息存在形式繁多,难以应用一般 CBR 技术解决。下面我们在引入一些基本概念的基础之上,对这些关键技术逐步展开讨论,并将产品模型理论和 CBR 技术紧密结合起来。

#### 3.1 基本概念

定义 1 设计实例 (Design Case) 是已有的满足一定设计需求的设计结果。

根据产品的层次可分为产品实例、机构体实例、装配体实例、零件实例和结构单元实例。

定义 2 实例动态视图 (Case Dynamic View, CDV) 设计实例产生过程中的一个状态,这个状态可以用一组特征属性和它们之间关系的集合表示。

定义 3 实例静态视图 (Case Static View, CSV) 实例动态视图中具有一定关联关系的特征属性集合。

实例动态视图和静态视图反映了实例的过程性和暂态性。实例的动态视图可以有多个静态视图。

定义 4 实例索引 (Case-Index, CID) 实例的标识,是一实例区别于另一实例的特征属性。

实例索引有全局索引、动态视图索引和静态视图索引之分,全局索引是实例所有动态视图的共性标识;动态视图索引 (Case Dynamic View Index, CDVI) 和静态视图索引 (Case Static View Index CSVI) 和全局索引具有相似的含义。

定义 5 实例原型 (Case Prototype, CP) 某一实例集中所有实例的共有属性的抽象表示,是实例类。相应地有实例原型动态视图 (Case Prototype Dynamic View, CPDV) 和实例原型静态视图定义 (Case Prototype Static View, CPSV)。

图 42.5-4 表示实例动态视图和静态视图、实例原型和实例库等的关系。

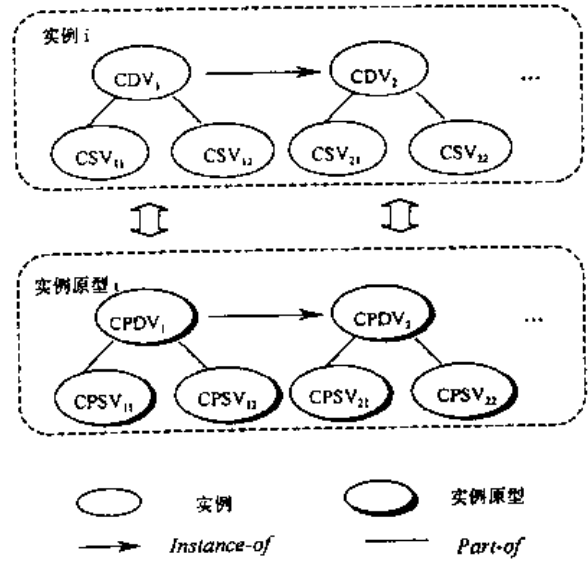


图 42.5-4 实例、实例视图和实例原型等的关系

联系到上面的有关实例、实例原型的定义,变型设计中存在以下基本概念:

1) 产品设计中,设计实例可以分为产品、机构体、装配体、零件和结构单元实例,这些实例之间的关系是整体和部分的关系,反映实例的结构组成关系。

2) 每个设计实例在理论上均有动静态视图定义。产品实例由反映产品模型的特征信息集定义,产品实例动态视图分为概念设计模型视图、装配设计模型视图和产品定义模型视图,每个动态视图反映了产品实例变化过程的一个状态,并由多个静态视图定义。机构体、装配体、零件及结构单元实例的动静态视图定义采用和产品实例相同的定义方式。设计实例的动态视图之间具有类和子类的关系,反映实例产生时间上的连续性和先后关系。

3) 实例原型也可以分为产品、机构体、装配体、零件和结构单元实例原型,它们之间同样具有整体和部分的关系。产品实例原型具有三个动态视图:概念设计实例原型、装配设计实例原型和产品定义实例原型。

我们在此定义的实例原型为一组实例的所有公有属性的抽象表达,实例原型和实例的关系就是对象类和对象的关系。在实例和实例原型中,仍将数据的表达和知识的表示分开。实例原型的产生是由归纳性知识决定的,它们存储在系统的知识库中。

一般认为实例原型具有归纳性知识,实例则不具有这些知识,知识并不作为实例的一部分存入实例库,实例只是实例原型的具体化,体现为属性、属性值的匹配,这说明在实例中没有记录设计历史。运用实例原型中的归纳性知识就形成实例,但条件不同则形成不同的实例,实例原型和实例均没有反映这些过程。

在实例中记录了归纳性知识的应用过程(通过记

录所使用的规则,形成确定的规则集)来记录设计不同实例的设计过程,这为实例的重用、修改奠定基础。参数设计不能解决所有设计问题,其原因就是在设计过程定性约束知识的记录、表达和处理的欠缺。

### 3.2 实例、实例原型的关联关系

在一个确定的工作领域中,实例之间存在着各种各样的关联关系,分析实例之间关联关系的目的,一方面既有利于实例的表示,同时也利于实例的组织和管理。

在前面我们定义了实例和实例的抽象类——实例原型,这反映了实例之间的一种关系,即一般抽象和具体对象的关系——分类关系(instance-of)。实例之间的关系不仅如此,尚有整体和部分之间的关系(part-of),如产品实例包含了机构体实例,机构体实例包含装配体实例、零件实例,装配体实例又包含了子装配体实例和零件实例等,这样就形成了一种分解层次关系,如图 42.5-5 所示。

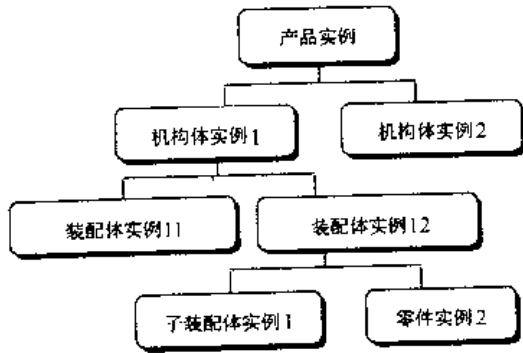


图 42.5-5 实例分解层次关系 (part-of)

实例原型也有着不同的抽象层次关系。抽象是人们求解设计问题的基本方法,设计原型便是对同类设计实例的抽象。鉴于设计问题的复杂性,我们对设计原型进行进一步分类和抽象。在实例原型库内定义不同的抽象层次,并在相关设计原型之间建立 kind-of 关系,构成抽象层次结构,如图 42.5-6 所示。在抽象层次结构树中,只有叶节点的设计原型才有对应的设计实例。

如图 42.5-7 所示为液压支架的实例原型抽象层次结构。

实例推理系统总是力图在抽象层次结构的最低层(设计实例)找到满足用户需求的问题求解方案,只有当找不到时,才向层次结构的上层回溯检索。如此进行,直至最终将某一抽象层次的设计原型作为参考设计方案返回用户。

因此,系统不仅能向用户提供具体的问题求解方案,以支持对设计实例的重用,还能通过抽象的实例原

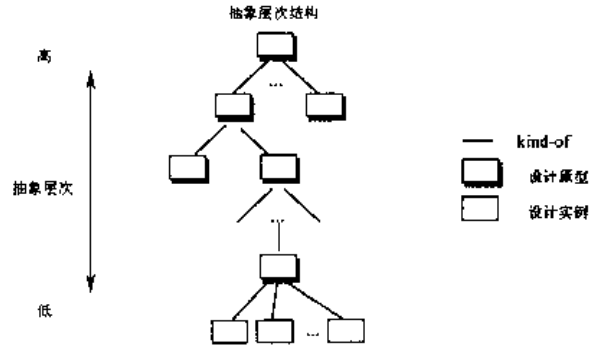


图 42.5-6 设计原型的抽象层次结构 (kind-of)

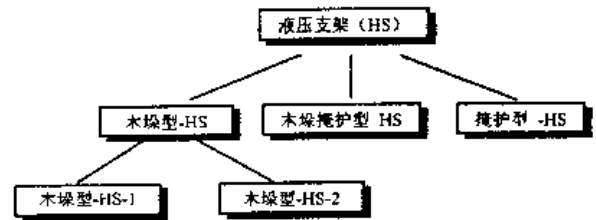


图 42.5-7 液压支架实例原型的抽象层次结构

型向用户提供问题求解的指导。

实例或实例原型与它们的内部动态视图之间是 part-of 关系,而这些动态视图之间则有类和子类之间的继承关系(kind-of)。

### 3.3 实例、实例原型的表示、索引和组织原理

#### 3.3.1 实例和实例原型表示

在实例推理系统中,实例的表示和索引、组织是密切相关的,且实例的表示和实例的内容也是相辅相成的,但又有区别。实例的表示和它的内容密切相关,但并不需要将所有这些内容均表示出来。实例的表示和实例的应用有关,例如要描述一个减速器产品,当描述的目的不同时,描述的内容也可不同。若向用户推荐产品,只需用产品名称、减速比、质量、最大转矩、额定输入和输出转速、价格等即可,并不一定要描述是几级传动、传动型式是什么等;但若是改造一个减速器,则描述这个减速器的应是另一组相关属性。

本文研究的应用于产品变型设计的实例推理系统中,我们表达一个设计实例的目的是为了当用户提出一个新设计需求时,能找到与此需求相匹配的相似产品。用户需求转化为设计需求模型时,形成了三个层次——概念设计需求、装配设计需求和产品定义需求,这是实例表示的第一个重要内容。

实例中必须表示的第二个方面内容是设计方案。检索和重用对应于不同设计需求的设计方案是实例推理的目的,而有效地表达设计方案是所有 CBD 系统必须关注的重要问题。

在设计过程中,设计人员的经验知识和设计知识的不断加入,使设计方案的最终形成又受到许多约束的限制,它们使实例间的区别具体化。实例的检索、重用必须考虑这些约束条件。这些约束为定性约束和定量约束,它们也必须成为实例表示的内容。

上面分析了实例应具有的内容,从这些内容我们可以了解一个实例的基本状况,并可以提取对应于实例的实例对象数据模型,以便对实例对象作进一步的分析和研究。但如何表示实例的内容,以利于实例的操作,这是CBD系统必须首先要解决的一个技术问题。

实例的表示方法和实例的内容相关,同时也关系到实例检索的效率、质量以及实例库的组织。在目前的CBD系统研究中,常用的实例表示方法有:特征属性-值对、文本、面向对象方法、图和多媒体表示等。用特

CasePrototype

```
{
    string CP-Name;           //实例原型的名称
    int CP-ID;               //实例原型标识
    sting CP-Type;          //实例原型类型:产品、机构体、装配体、零件和结构单元实例原型
    Boolean Verification=true; //实例原型修改性标识,为true表示可以修改
    Design-Requirement-Model; //实例原型的用户需求模型
    CP-DV CPDV-List;        //实例原型的动态视图原型集
    Relation Part-of (value CPID); //本实例原型是标识为CPID的实例原型的部分原型
    Relation Kind-of (value CPID); //本实例原型是标识为CPID的实例原型的子原型
    RuleSet;                //设计定性约束集合。
}
```

实例原型的类型CP Type是标识实例原型、产品实例原型、机构体实例原型、装配体实例原型或是零件实例原型、结构单元实例原型。

实例原型的动态视图原型集CPDV-List给出了该实例原型具有的动态视图的集合,本文只研究产品实例原型具有的动态视图。

设计实例原型的可修改性标识表示实例是否可以被修改,形成一个新的设计实例,有的实例原型是明确不可修改的,如企业中的产品模块,只能被选择,不可被修改。

CP DV

```
{
    string CPDVName;        //实例原型动态视图名称
    int CPD_ID;            //实例原型动态视图标识
    Boolean Verification=true; //实例原型动态视图修改性标识,为true表示可以修改
    Design-Requirement-Model; //实例原型的用户需求模型
    CPDVFeature-List;      //实例原型动态视图特征属性集合
    CPSV-List;             //实例原型动态视图的静态视图集
    RuleSet;               //设计定性约束集合
}
```

征属性-值对表示实例,简单且操作方便,被广泛采用为实例的基本表示方法。但若只用这一种表示方法,则无法反映实例之间的各种关系,故一般必须辅助其他方法来组织这些特征属性,如FBS(Function-Behavior-Structure)模型、相关性约束网络和定性模型等设计模型的应用。

我们在此应用面向对象技术表示实例,将实例原型定义为对象类,实例作为实例原型的对象。特征属性是实例原型和实例的基本表示方法,同时应用了文本、图形等表示方法以表达不同型式的实例内容。设计实例的需求模型和实例的动态视图和静态视图主要采用特征属性-值对的集合表示。

综合上述因素,我们将实例原型用如下面向对象方法表示:

关系集合,包含上文所述的反映实例原型的分解层次结构的part-of关系和反映抽象层次结构的kind-of关系。

定性约束关系集合是产生此实例原型的定性知识的表示,一般以规则序号集的形式出现,所有规则存于全局知识库中。

实例原型的动态视图表示实例原型的产生过程中的一个状态,分析或应用一个实例原型,有时必须利用其动态视图,因而有必要表示这些视图。实例原型的动态视图和实例原型的表示基本相同,定义如下:

实例原型动态视图的静态视图采用和动态视图基本相同的表示方法。在实例原型或其动态视图的表示中,设计特征属性集是主要内容。而对于每一个设计特征属性,除了特征名之外,还要说明该特征的类型(TYPE)、默认取值(DEFAULT)、单位(UNIT)和

Case

```
{
    string C-Name;           //实例的名称
    int    C ID,           //实例标识
    sting C Type,         //实例的类型:产品、机构体、装配体、零件和结构单元实例
    Boolean Verification= true; //实例修改性标识,为 true 表示可以修改
    Design-Requirement-Model; //实例的用户需求模型
    C-DV CDV List,       //实例的动态视图集
    Relation Part-of (value CID); //本实例是标识为 CID 的实例的部分实例
    Relation Instance-of (value CPID); //本实例是标识为 CPID 的实例原型的实例
    RuleSet;             //设计定性约束集合
}
```

实例的动态视图表示实例的形成过程中的一个状态,实例的重用和修改必然涉及过程因素,因而有必要

C-DV

```
{
    string CDVName;       //实例动态视图名称
    int    CDV ID,       //实例动态视图标识
    Boolean Verification= true; //实例动态视图修改性标识,为 true 表示可以修改
    Design Requirement-Model; //实例动态视图的用户需求模型
    CPDVFeature-List;    //实例动态视图特征属性-值对集合
    CPSV-List;          //实例动态视图的静态视图集
    RuleSet;           //设计定性约束集合
}
```

实例动态视图的静态视图采用和动态视图基本相同的表示方法。实例及其动态视图中的特征属性是对实例原型中的特征属性的赋值。

在实例原型、实例及它们的视图中,定义的用户需求模型采用有关特征属性集合表示。

设计实例库可看成是一组相互间有继承关系的层次体系,其中设计实例原型位于高层而设计实例位于低层。每一个设计实例均通过一个 instance-of 关系与相应的设计实例原型相连。一个设计实例可分解为一组子设计实例,并通过 part-of 关系与其相连。

### 3.3.2 实例和实例原型的索引、组织

实例、实例原型的组织和索引机制、实例提取算法及实例的管理是密切相关的,良好的实例组织形式不仅有利于实例的检索,也有利于实例的管理。

实例的组织结构常用的有:

1)平直结构 这种结构采用线性数据结构存储指

取值范围(RANGE)。

设计实例包含了满足一个特定设计需求的所有设计结果,它是相关实例原型的实例化,是在满足一定约束条件的情况下,对实例原型中各特征属性赋值的结果。实例的表示如下:

表示这些视图。实例的动态视图定义如下:

向每个实例的指针,实现对实例的组织。它用实例的名称作为实例的索引,实例的检索是对结构中的每个元素的搜寻,并和目前的待解决问题相比较,因而不会遗漏任何相似实例。实例库的管理方便,但检索的效率较低,只适用于设计实例较少的情况。

2)基于特征的实例库结构 对某一领域中的所有实例,按预先确定的一组特征属性存储这些实例,每个特征属性指向具有此特征属性的实例标识,而将这组特征属性存储在一个线性结构中。实例检索时,只有符合设计需求的实例方被检索,方法是特征属性-值匹配。这种方法中,确定划分实例库的特征属性集是关键。特征属性可以是抽象的对象,也可以是具体的实体,可以是确定的概念,也可以是模糊的现象。它只要求特征属性可以有效地描述实例之间的相似性,但这种方法中的实例组织结构是静态的,无法支持实例的灵活检索机制。

3)分类层次结构 这种结构是将索引特征属性组

织成一个层次结构,利用从一般抽象特征到相对具体特征,将实例分类组织在层次结构的叶节点中。这种结构利于实例的渐进检索,检索效率高于前面两种方法,因为它只考虑所有实例的一个子集;但这种组织方法使实例的管理较为复杂,对非线性数据结构的操作的难度和效率远远低于线性结构。

在本文讨论的实例库中,存在两类实例:实例原型和实例,但实例原型是实例的抽象,是一类实例的集合的公共属性的抽象,因而实例原型本身就是对一类实例的一种索引,并且实例和实例原型都有动态视图的分类,它们的动态视图之间也一样具有抽象和具体的关系。

考虑到这些因素,本文的实例组织采用两层组织策略:基于特征的实例原型组织和基于实例原型的实例层次组织形式。这就是在实例原型的组织中,采用线性结构,因为首先进行的是实例原型的检索,然后根据实例原型的分解层次关系(part-of)及和实例之间的instance-of关系,确定属于特定原型的实例或实例视图,如图42.5-8所示。

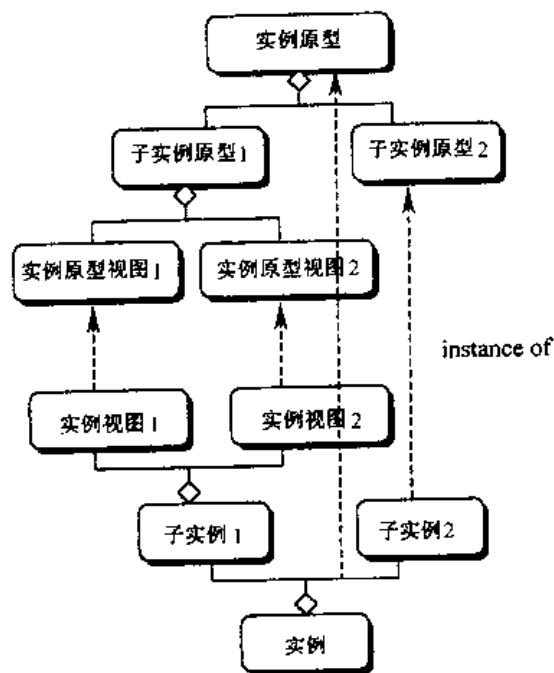


图 42.5-8 实例库的组织

在基于特征的实例原型的实例库组织中,不同种类的实例原型具有不同性质和不同数量的特征属性。为了提高特征的近似匹配的效率,我们将这些特征归类,形成实例的四类特征索引:

●实例原型名称 这样可以直接由实例原型名搜索到实例原型的完整表达。

●概念设计特征索引 具有相同概念设计模型的实例原型集。相同概念设计模型指功能相同、结构映射

视图相同,而设计参数属于一定的范围。

●装配设计特征索引 具有相同的装配设计模型的实例原型集。相同的装配设计模型指产品结构组成、工作原理相同,但装配体内部设计参数有可能不同。

●产品定义设计特征索引 结构相同、功能相同、结构组成相同及部件的结构形式相同但设计参数不同的一类实例原型。

实例的组织,采取分散组织方法,即一个产品实例的所有子实例不是集中地存在一个位置,而是分散存储在不同的实例库中,同时建立实例间的兄弟关系和上下层父子关系,形成一个对应于实例原型的实例分解层次关联树。

实例的分散存储和实例的分解粒度有关,实例分解的粒度过粗不利于子实例的应用,但若实例分解太细,又不利于实例的检索和存储。在处理这个问题上,我们采用的策略是,若一实例虽可再分解为子实例,但这个实例已被看作是可以利用的、较为固定的基本单元,这样的实例或子实例将不再实行分散存储。在本章中这种实例基本单元为装配体实例,不再研究它们内部的零件。

### 3.3.3 实例的检索

在实例依据一定的索引机制组织存储后,当面对一个新的设计问题时,就可以利用新问题描述进入实例库进行实例检索。实例检索的过程就是实例间相似性的判别过程和相似实例的提取过程。

实例的相似性判别有很多方法,它与实例表示、索引组织密切相关。

#### (1) 实例相似性判别条件

首先我们要定义两个实例的相似性条件,我们认为符合以下两个条件之一的就可以认为两个实例相似。

条件1 具有相同或相似设计任务的两个实例是相似实例。

条件2 具有相同或相似动态视图的两个实例是相似实例。

问题的关键是如何判别两个实例具有相似的设计任务或相似的动态视图。

对设计任务相同或相似的两实例是相似实例这一点是无可置疑的,但通过比较设计任务来确定实例的相似性会造成获得的相似实例太多这一无法避免的缺陷。因为设计任务只是产品的一种粗略描述,是产品用户和设计人员之间关于产品的基本性能的共识,主要反映用户对产品的需求,这样的需求只是产品设计需求的一个子集。在产品的开发过程中,设计任务是不断变化的、不断得到丰富的,设计需求也是不断地增加

的,这样的增加不断地缩小产品可用空间,最终的产品是产品空间中得到优化的一个点。因此产品的设计任务决不能成为相似设计实例判别的唯一依据,它只是一个起点,相似判别条件 2 就是以条件 1 为基础的。

上面说明了这样的一个事实:具有相似设计任务的两实例有相似性,但相似程度可能差别较大,解决这个问题唯一的办法就是对获得的相似设计实例进行实例动态视图的相似性判别,只有具有相同或相似动态视图的两实例才是最终的相似设计实例。

实例具有动态视图,而动态视图又具有静态视图,因此对两个实例,其相似性应从它们的动态视图和各个动态视图包含的静态视图各方面来考虑。

在产品变型设计中,实例动态视图的数量和性质是随实例类型变化的,但一般对变型设计影响最大的是产品。为了不失一般性,在论述实例相似性问题时,假定实例都具有三个动态视图,即概念设计动态视图、装配设计动态视图和产品定义动态视图。

实例的相似性从动态视图判别是基于这样的事实出发的:

- 两个实例的概念设计动态视图相同或相似,但装配设计动态视图或产品定义动态视图很可能不相似;具有相似装配设计动态视图的两个实例也存在这种问题。

- 即使是同一动态视图其各个静态视图也有相似性问题,即两个动态视图的相似性必须从其各自包含的静态视图的相似性来判别。

目前 CBR 推理中很多实例相似性判别方法,均不适应目前这种状况,“Fish and Sink”算法中虽有“视图”的概念,但这些视图之间没有时间约束关系,且它们自身也没有包含更深层的静态视图概念。

实例的相似性判别中还存在一个问题,即实例的相似性是随用户应用实例的不同目的而不断变化的,在一定条件下相似的实例,在条件发生变化后,可能就成为完全不相似的实例。这种动态变化状况和实例的静态组织形成了矛盾。

实例的动态视图和静态视图描述了实例的动、静态特征属性集,这些特性是已设计完成的实例的固有特性,反映满足用户过去特定需求的结果。目前实例库的一般组织方式是由特征属性确定的,这种组织方法反映了实例特征属性的确定的重要性程度,这种重要性程度是满足用户过去的需求,因而也就固定了特征属性之间的重要性程度关系。这些与用户实际需求是矛盾的;不同的用户在不同的时间,要求用不同的特征属性集和不同的特征重要性程度来反映实例的特性,因而特征属性集和它们的重要性程度是不断变化的。

因此,反映实例特性的特征属性集是随用户的不同

同目标而动态变化的,而且这些特征属性的重要性要求也是动态变化的。这要求实例库的组织必须是动态的,具有较大的灵活性,但这和要求具有良好的实例库结构,以利实例高效搜索和管理是明显矛盾的。

总之,我们在实例动态视图的相似性判别中,必须考虑以下问题:

- 1) 在实例的相似性评判中,要考虑实例的动态视图和动态视图中的静态视图之间的相似性;

- 2) 实例动态、静态视图的特征属性集、特征属性重要性程度是随用户需求动态变化的,即用户的不同目标一般会导致实例间动态视图和动态视图中静态视图的相似性发生变化。

在分析了实例相似性判据之后,可得出实例及实例原型的相似性判别步骤如下:

STEP1: 利用“相似判别条件 1”判别两个实例的设计任务的相似性,若设计任务不相似,则无需继续进行,否则继续;

STEP2: 利用“相似判别条件 2”继续进行实例的动态视图相似性比较,得到的相似实例即为满足要求的相似实例,这个过程是进行动态视图的每个静态视图相似性判别,并在判别过程中必须充分考察特征属性集的变化和特征重要性程度的变化。

每个视图的相似性比较基本上采用特征属性值匹配和约束规则匹配方法,使定性约束和定量约束都得到满足。

## (2) 基于索引拓展的实例检索模型

实例的检索是从用户的产品设计需求开始的。在产品变型设计中,产品的设计任务在由用户提出后,经过设计人员的加工整理,形成产品设计需求模型。这个产品设计需求模型将用户需求渐进分解为产品概念设计需求模型、装配设计需求模型和产品定义需求模型。这样分解的目的和意义是十分明显的,因为产品的设计需求是一个随设计过程渐进增加的过程,在产品设计的不同阶段,设计人员会根据用户的不明确的需求,增加满足这个设计阶段要求的新的设计需求,如技术参数、技术属性等。

设计人员根据产品设计需求模型应用规则推理技术,定义一个粗略的初始产品动态模型:概念设计源模型、装配设计源模型和产品定义源模型,它们组成源实例。

实例的检索开始于以源实例检索实例原型,由相似实例原型确定相似实例,这个过程按以下步骤进行:

STEP1: 由概念设计源模型(ICP1)从实例原型库中检索相似实例原型,检索出的实例原型形成一个具有相似概念设计模型的实例原型集 CPS。

STEP2: 依据装配设计源模型(ICP2)进行相似

实例原型检索,进一步缩小 STEP1 获得的相似实例原型空间 CPS,形成具有相似装配设计模型的相似实例原型集 APS;

STEP3:在 APS 集合中,依据产品定义源模型(ICP3)检索具有相似产品定义模型的实例原型,形成实例原型集 DPS,并从中提取最相似的实例原型(Ms-DPS),进行 STEP4,若没有相似产品定义实例原型,则检索回溯至 STEP2,这时必须修改装配设计源模型;若这一层的检索也失败,则检索回溯至 STEP1,相应地修改概念设计源模型进行实例原型的重新检索。

STEP4:从与相似实例原型对应的实例集合中提取最相似的实例,相似实例检索完成。

实例检索的结果是三个层次的相似实例集及最终可提交实例修改的最相似实例。

从以上这个过程可以看出,实例检索是依据产品设计需求模型的三个子模型和产品的三个设计源模型,由简单到详细和由详细到简单循环往复进行的,这种检索模型不同于一般的实例检索模型,我们称之为基于索引拓展的实例检索模型。图 42-5-9 描述了这个检索模型。

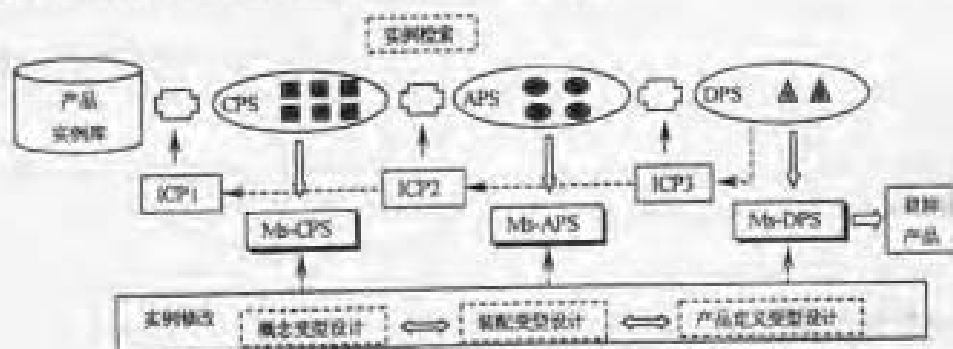


图 42-5-9 基于索引拓展的实例检索模型

基于索引拓展的实例检索模型中的“索引拓展”表现在两个方面,一是实例索引的拓展,设计人员首先根据用户设计需求拓展实例的索引,形成了和三个设计阶段对应的不同的设计需求模型和实例源模型;二是实例检索空间的不断拓展,实例的检索虽然开始于具有相似概念设计模型的实例检索,但相似实例的提取并不是首先在此阶段,而是开始于设计的较详细阶段——产品定义阶段,这时的实例检索空间较小,检索效率较高,且因为首先进行了实例索引的拓展,形成了和低层设计阶段对应的设计需求,使这种检索的质量也有了保证;若检索失败,则回溯至上层设计过程中检索相似实例,使检索的空间得到了拓展,若直接根据用户需求进行实例检索,则检索空间很大,检索的效率低,检索得到的实例的质量也较差,因为设计需求有待补充和完善,这样检索的实例不能完全反映用户的设计需求。

Kolodner 也曾经提出过一种基于索引拓展的实例检索模型,它是通过增加用户初始设计需求中没有包括的分类特征或其他关键特征来拓展用户设计需求,如图 42-5-10 所示,它的实现是借助于设计模型或原型,直接产生一个拓展后的新的设计需求,增加的特征属性是由设计模型提供的,因而其中有根据初始设计需求寻找可提供新设计特征的设计模型的过程,而这个过程中选择增加何类特征较为盲目,且和设计阶段没有关系。

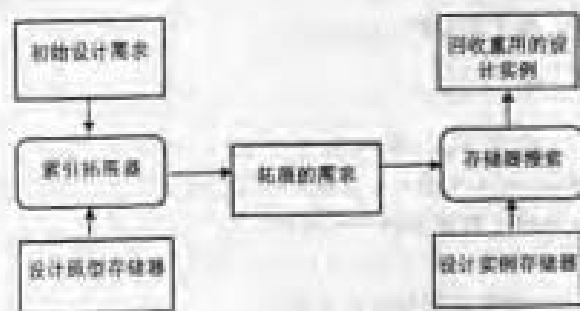


图 42-5-10 采用加工索引的实例检索过程

(3) 实例相似性判别算法

实例的相似性判别条件和实例检索模型给出了实例的相似性判别步骤和实例检索的过程模型,但仍然没有完全解决实例的检索问题,两实例的相似达到什么程度才可以作为满足需求的相似实例提取出来,如何搜索实例库,是全局搜索,还是局部搜索,实例检索的终止条件等问题仍需解决,这些就是实例相似性判别算法要解决的问题。

为了讨论的方便,我们引入以下定义,并且所讨论的算法对实例,实例的动态视图和实例原型均有效。

设  $\Omega$  是研究领域实例集合,  $C_1$  和  $C_2$  是进行相似性比较的两个实例,  $V_1$  和  $V_2$  是相比较的两个实例的对应动态视图(但当比较的是实例的两个动态视图时,以下定义中的视图  $V_1$  和  $V_2$  则是指实例动态视图的静态视图)。

定义 7 距离 (Distance) 是定义在集合  $\Omega$  上的一个函数  $\delta: \Omega \times \Omega \rightarrow [0, 1]$ , 且  $\delta(C_x, C_y) > 0$ , 而  $\delta(C_x, C_x) = 0$ , 是实例相似程度的表达。

定义 8 视图定义距离 (Case View Defined Distance, CVDD) 根据预定义的评价标准定义的存储于实例库中的两实例对应视图之间的相似距离, 且  $\delta_d(V_{x_i}, V_{y_i}) > 0$ , 而  $\delta_d(V_{x_i}, V_{x_i}) = 0$ 。

定义 9 实例定义距离 (Case Defined Distance, CDD) 根据预定义的评价标准、视图定义距离及视图权重系数, 定义的存储于实例库中的两实例之间的相似距离, 且  $\delta_d(C_x, C_y) > 0$ , 而  $\delta_d(C_x, C_x) = 0$ 。一般有:

$$\delta_d(C_x, C_y) = \sum_{i=1}^n W_{ci} \delta_d(V_{xi}, V_{yi})$$

其中  $W_c$  是实例库中实例视图的定义权重向量, 它反映组成实例的各个视图的重要性程度差别。

$$W_c = (w_1, w_2, w_3, \dots, w_n), \sum_{i=1}^n w_i = 1$$

定义 10 视图用户权重向量 (User Case View Weight Vector, UCVWV) 用户定义的组成实例的各个视图的权重系数组成的向量。当  $w_{ui} > 0$  时表明视图  $V_i$  属于源实例。向量如下:

$$W_u = (w_{u1}, w_{u2}, w_{u3}, \dots, w_{un}), \sum_{i=1}^n w_{ui} = 1$$

定义 11 实例应用距离 (Case Application Distance, CApD) 在相同的实例视图用户权重向量  $W_u$  作用下的两实例之间的实际相似距离。

$$\delta_s(C_x, C_y, W_u): \Omega \times \Omega \times W_u \rightarrow [0, 1]$$

一般定义实例应用距离为

$$\delta_s(C_x, C_y, W_u) = \sum_{i=1}^n w_{ui} \delta_d(V_{xi}, V_{yi})$$

定义 12 试验实例 (Test Case, TC) 每次实例检索中选定的和源实例作相似性判别的实例。

定义 13 试验距离 (Test Case Distance, TCD) 试验实例和源实例之间的实例应用距离。

定义 14 近邻实例 (Case Neighbor, CN) 在实例库中和试验实例具有用户设定相似距离的实例。

定义 15 近邻距离 (Case Neighbor Distance, CND) 试验实例和其近邻实例之间的实例应用距离。

实例的相似性判别和检索的前提有二, 一是源实例的视图定义和与其比较的试验实例的对应视图反映的内容相同, 所不同的是源实例中各个视图的权重系数和试验实例不同, 这种权重系数反映了用户对一个实例的特殊要求; 二是实例库中存储实例之间的相似性, 即实例定义距离和实例视图定义距离。这种相似性

是满足一定预定义条件的, 反映了实例之间的基本相似性质, 可以认为是实例应用历史的反映。

实例的相似性判别随着实例检索的进行而渐进进行, 检索的目的是快速寻找和用户需求相适应的源实例的相似实例。方法是通过和实例库中的已有实例进行相似性比较, 在确定了一个相似性标准后 (比如实例应用距离小于 0.5), 就可以从实例库中检索满足特定需求的相似实例。这样求得的相似实例是一个集合, 因此就有最优实例的判别问题。

在实例库规模较小的情况下, 可以通过实例库的全局搜索, 匹配实例库中的每一个实例, 找出相似实例, 并选取最优实例; 但若实例库的规模很大, 这种全局搜索方法很明显效率太低, 因此必须研究高效高质量的实例检索算法。

为了解决这个问题, CBR 的研究者们作了不少努力, 提出不少行之有效的方法, 如 Goos 的爬山法、KRITIK 的功能索引、SEED 系统的层次匹配和 CADET 的图匹配、BurkHard 的 CRNS (Case Retrieval Nets, 实例回收网络) 等。这些方法一般都是利用预定义的实例特征属性相似关系作为相似性判别的基础, 并根据实例的应用, 定义特征的权重系数, 以区分特征对实例的重要性, 但通常以少数特征属性很难描述实例的性能和状态, 且有的实例相似性难以用特征属性或特征属性值来衡量。这在产品设计领域是经常面临的问题, 如两个装配体, 它的组成零件完全相同, 但当零件的装配结构关系不同时, 它们就定义了不同的功能和行为, 因而就不是两个相同或相似的实例, 而结构关系是很难用特征属性等来衡量的。因此在设计领域应用实例推理技术必须研究合适的算法解决这些问题, Febal 是应用于建筑领域中一个实例推理系统。其中采用了“Fish and Sink”方法进行实例的相似性判别, 通过引入“视图”概念, 使实例的检索在用户意图控制之下进行, 使视图之间的相似性处于动态变化之中, 但这些视图是实例的不相关的侧面, 反映了用户的不同需求。

本节在上述有关定义的基础之上, 提出了“实例多视图检索算法”, 这里的“视图”是实例的一个动态视图 (或动态视图的一个静态视图), 反映了实例 (或实例动态视图) 的一个状态, 与实例的过程性和动态性相关。以下三个假设是基础。

假设 1 若源实例和试验实例之间的实例应用距离较大, 则此试验实例的近邻实例也就具有很少的可用性。

假设 2 源实例和试验实例的近邻实例之间的实例应用距离不小于试验距离和近邻距离之差; 用公式表示为



$$\delta_a(A, V) \geq \delta_a(A, T) - \delta_a(T, V)$$

式中  $A$ ——源实例;

$T$ ——试验实例;

$V$ ——试验实例的近邻实例;

$\delta_a(A, V)$ ——源实例和近邻实例之间的实例应用距离;

$\delta_a(A, T)$ ——试验距离;

$\delta_a(T, V)$ ——近邻距离;

**假设 3** 源实例和试验实例的近邻实例之间的实例应用距离不大于试验距离和近邻距离之和; 用公式表示为

$$\delta_a(A, V) \leq \delta_a(A, T) + \delta_a(T, V)$$

这样就定义了试验实例的每个近邻实例与源实例之间的实例应用距离的可能范围。每一个不同的试验实例都对其近邻实例和源实例之间的距离作这样的操

作, 最终可以确定和源实例之间有一定的实例应用距离的相似实例集及它们按相似距离的排序。这个过程就是相似实例的检索过程。

在以下算法中, 为每个试验实例的近邻实例保存了最大的  $V.\text{mindist}$  和最小的  $V.\text{maxdist}$ , 而区间  $[V.\text{mindist} \ V.\text{maxdist}]$  定义了每个实例和源实例之间的可能相似距离区间,  $V.\text{maxdist}$  只能下降, 而  $V.\text{mindist}$  只能上升。这个区间是通过源实例和试验实例的一次比较获得的。通过有限次的采集试验实例, 不断对这个区间的上下限进行操作, 最后得出试验实例的近邻实例和源实例之间的实例应用距离存在的区间或精确的实例应用距离。这样就可以判别试验实例的近邻实例与源实例的相似性, 而不需再直接对每个近邻实例进行相似性判别, 提高了检索的效率。用算法描述如下:

- 1) Let CB be a list of all cases in the case base.
- 2) For each case  $F \in CB$  let  $F.\text{mindist} := 0$  and  $F.\text{maxdist} := 1$ ;
- 3) While not OK and not interrupted
  - (a) move precision line PL;
  - (b) choose a case  $T$  with  $T.\text{mindist}$  on PL;
  - (c)  $T.\text{mindist} := \delta_a(A, T, W_a)$ ;
  - (d)  $T.\text{maxdist} := T.\text{mindist}$ ;
  - (e)  $\forall$  case  $V$  are the Case Neighbors of case  $T$  with  $V.\text{mindist} \neq V.\text{maxdist}$  do
    - I.  $V.\text{mindist} = \max(\delta_a(A, T) - \delta_a(T, V), V.\text{mindist})$ ;
    - II.  $V.\text{maxdist} = \min(\delta_a(A, T) + \delta_a(T, V), V.\text{maxdist})$ ;

在上述算法中,  $\text{mindist}$  和  $\text{maxdist}$  表示每个实例的最小相似距离和最大相似距离,  $A$  表示源实例, OK 表示用户需求得到满足的条件, 如提供实例应用距离为某一阈值  $S$  的所有相似实例; 而 PL 是由实例库中暂时不满足要求的实例的最小相似距离  $\text{mindist}$  所确定的相似距离标识线, 此线随着检索过程的前进而不断地变化, 并由此线确定下一次相似判别用的试验实例, 即所有  $\text{mindist}$  位于 PL 线上的实例将是下一次相似判别用的试验实例。若相似检索达到用户的要求, 则此时所有  $\text{mindist}$  和  $\text{maxdist}$  都越过 PL 线的实例将被提交给用户。

利用这个算法可以向用户提供实例应用距离为某一阈值  $S$  的所有相似实例, 或  $K$  个未排序的相似实例及  $K$  个经过排序的相似实例。我们将这些情况分别表示于图 42.5-11a、b、c 中。

图 a 表示给定一个阈值  $S$ , 凡  $\text{maxdist}$  低于  $S$  的实例 (图中的实例 a) 不需再直接判别, 它是源实例的相似实例, 而  $\text{mindist}$  大于  $S$  的实例 (如图中的实例 c) 也

就没有判别的必要。图 b 表示  $K$  个没有排序实例的  $\text{maxdist}$  低于 PL 线, 它们是源实例的相似实例。图 c 表示  $K$  个排序实例的  $\text{maxdist}$  低于 PL 线, 它们是源实例的相似实例。

在这个算法中, 关键是视图定义距离和视图用户权重向量的确定, 视图用户权重向量的确定和用户的设计需求有关, 而确定视图定义距离则可以根据特征属性值对的匹配情况, 特征属性匹配的越多, 则两个视图的视图定义距离越小, 两个视图越相似。这些工作可以在建立实例库时解决。

### 3.4 实例的修改

经过相似判别提取出的实例往往不可能直接满足新设计要求, 必须经过修改方可满足全部设计要求。

相似实例的修改是一个复杂的过程, 新的设计要求往往打破相似实例中的很多约束, 包括定性和定量的约束, 必须识别这些冲突约束方能顺利完成实例的修改。

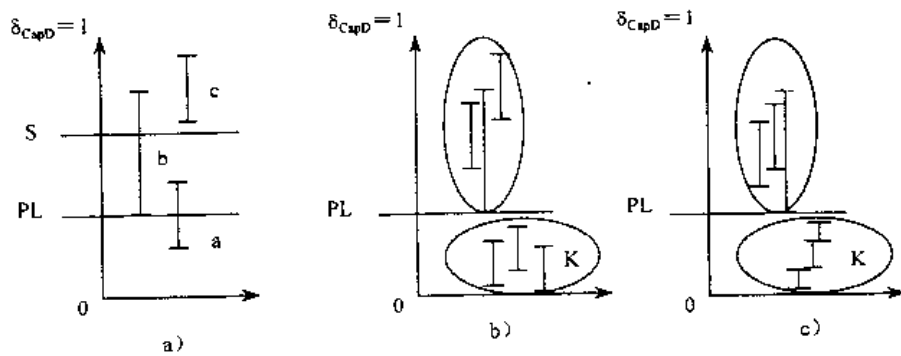


图 42.5-11 实例的相似距离

### 3.4.1 实例修改的基本问题和修改策略

实例的修改中存在三个基本问题,即定性冲突约束、定量冲突约束的处理和实例的分解组合问题。

定性冲突约束是首先必须解决的冲突问题。定性约束表示设计过程中的各种规则知识,因此定性约束的违反引起的冲突,可以由规则推理完成。我们在前面实例的表示中,每个实例的定性约束出一组规则集表达,冲突约束表现在对这些规则中的一些规则的违反,经过规则推理,可以完全消除冲突。

定量约束的处理是一个设计变量约束网络的求解过程。我们在设计建模时,已建立了产品、机构体及装配体等的设计变量间的约束关系图。确定冲突的特征属性变量,即可确定冲突的定量约束,用冲突发生后的新特征属性变量值,取代约束网络中的原变量值,进行约束网络的重新求解,最终消除冲突约束,这是一个典型的约束满足问题(Constraint Satisfaction Problem, CSP)。在解决这个问题的时候,必须区别特征属性变量的性质,并将特征属性变量从合理的求解空间中划分出来。

特征属性变量是实例特性的反映,在实例的不同动态视图和不同静态视图中,特征属性变量具有不同的性质和不同的表现。特征属性变量具有以下性质:

1) 层次属性 上层的特征属性变量在下层无修改的权限,这反映了特征属性变量的继承关系;变量层次有概念设计层、装配设计层和产品定义层三个层次。

2) 可修改性 有些特征属性变量在设计过程中是可修改变量,而有些则是不可修改的变量,不论它处于何种变量层次;

3) 在同一层次,特征属性变量的性质有可能是不同的,同层特征变量可以分为关联变量(或称为层际属性变量)和独立变量(或称为层内属性变量),关联变量反映了对象间的联系,独立变量不可影响关联变量,但关联变量则可决定独立变量。

在规定了特征属性变量的特性后,设计约束网络的求解很明显不同于一般的 CSP 问题,整个设计约束网络的求解转化为多个约束网络求解问题,如图 42.5-12 所示。

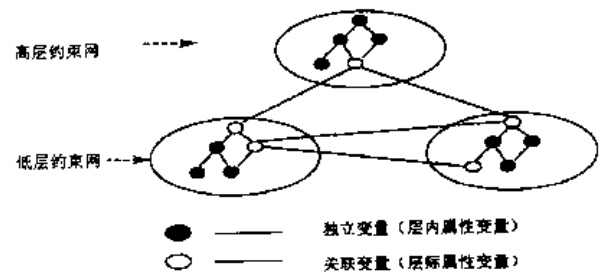


图 42.5-12 约束网络层次图

在图 42.5-12 中各个子约束网中的关联变量形成了一个关联变量约束网,约束求解必须首先求解这个约束网,然后求解关联变量发生变化的子约束网,在子约束网络的求解过程中,这些关联变量是不可修改的。

这种 CSP 方法不同于一般的 CSP 方法,考虑了实际存在的产品设计规划,而不是盲目地确定求解规划,将设计约束网络的求解简单地看作是一个求解非线性方程组的问题。采用这个策略,虽然最终的约束网络求解也是和普通的 CSP 方法相同,但问题求解的规模和速度已不可同日而语! 关联变量和独立变量的划分是这种方法的关键,一般约束求解系统中,也有约束规约的问题,求解的思路和这种变量划分相似,但这里的划分是有现实依据的,是和主观设计意图相关的,和一般的约束规约不同。例如,在产品设计中,经常有产品模块重用的情况,这时这个产品模块的设计参数网络中关联变量和独立变量都不可以变化。因此关联变量的可修改性是经常变化的,与产品资源管理及设计领域等均相关。

实例修改中必须注意的第三个问题是实例的分解和组合。实例的修改与实例的分解、组合是密切相关的。在我们提取出一个相似实例后,修改这个实例以满足目前设计要求的关键是冲突约束的识别。在识别冲突约束过程中,有时必须分解相似实例,对它的结构组

成进行深入地剖析,分析冲突约束的真正所在;不满足要求的相似实例,它不符合要求的部分很有可能存在于它的某一个组成单元中,对这个组成单元的修改比对整个实例的修改要容易的多。有实例的分解,就有实例的组合。对分解后的实例单元进行修改之后,就必须进行实例单元的组合,这个过程在产品实例的修改过程中是经常发生的事情。模块化设计中的模块组合与此就十分相似。实例组合过程中存在的约束冲突问题又是定性约束冲突和定量约束冲突问题,这些问题的解决可以采用前面的策略。

### 3.4.2 基于约束满足技术的实例修改过程

由上面的描述可知,实例修改操作中的关键技术是约束满足技术:实例检索提供了一个不一致的约束满足问题作为初始解,实例修改是约束满足问题的一致性检查和求解。

基于约束满足技术的实例修改过程可分为以下五个步骤:冲突约束的检测,冲突变量类型的确定,冲突变量的选择,冲突变量的修改,变量修改后的影响传播。如图 42.5-13 所示,这是一个循环迭代过程,实例修改不断检测出新的冲突约束,产生一组新的变量赋值,直至最终求得约束系统的一致解。若不存在一致解,则要求实例修改产生一个最优解作为进一步修改的基础。以下对各步骤作一个简要说明。

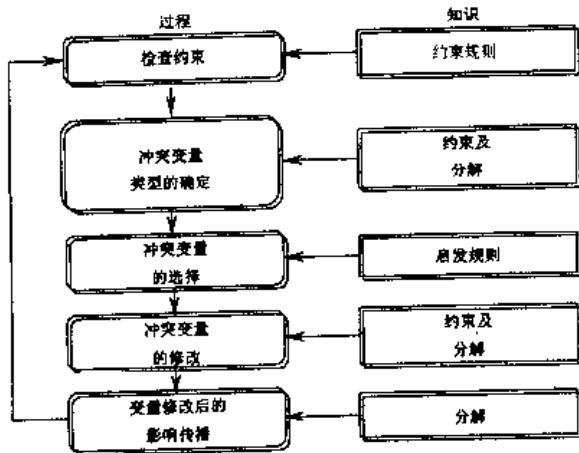


图 42.5-13 基于约束满足技术的实例修改过程

**冲突约束的检测:**实例检索生成的(子)设计实例提供了一组设计特征属性变量赋值,设计原型记录了各变量之间的一组约束关系,冲突约束检测的任务是发现不满足的约束关系,并找出所有的冲突变量。

**冲突变量类型的确定:**将冲突变量按定义分为关联变量、独立变量和它们所处的层次,首先确定冲突关联变量约束网络,作为实例修改操作的对象。若没有冲突的关联变量,只有独立变量存在冲突,则只是一个局

部约束网络的求解问题。

**冲突变量的选择:**根据冲突关联变量所在的层次,确定冲突变量的修改次序,若原来它处于高层约束网络中,则它具有优先的修改顺序。

**冲突变量的修改:**在冲突变量的值域中选择一个新的赋值。变量值选择的基本要求是:赋值后使与该变量相关的所有约束尽可能一致,即满足局部一致性要求。变量值选择的进一步要求是:赋值后使整个关联变量约束网络的冲突约束数量尽可能少,即满足全局一致性要求。

**变量修改后的影响传播:**冲突关联变量约束网络的求解是实例修改的关键,但不是实例修改的结束,它只解决了约束网络层间的关联变量的一致性求解,而在约束网络层内的各个变量之间又会因关联变量的修改导致不一致性,这是变量修改后的影响传播,引起新的冲突变量约束求解。但这种约束求解中,关联变量是不可改变的变量,否则会导致约束求解系统的过多往复循环,使约束系统求解效率低下。

虽然约束满足问题是人工智能中的传统问题,但在以上实例修改中有其特殊性。首先,实例修改中的约束满足问题是一个从初始的不一致解,经过不断修正直至生成一个一致解的变换过程,而不是传统的从无到有的构造过程,因而约束系统的求解宜采用修正法,而非传统的构造法。其次,由于在设计过程中设计约束系统常处于不完备状态,即存在过约束、矛盾约束或欠约束等情况,无法求得完整的一致解,因而实例修改中的约束满足问题是一个部分约束满足问题(Partial Constraint Satisfaction Problem, PCSP),即只在一定范围内存在含有不一致约束的最优解。本节对此问题不作深入讨论,有关此问题的深入研究可参见其他文献。

## 4 基于 CBR 的产品快速响应变型设计

为了提高新产品的开发速度和产品的质量,面对一个新的设计需求,我们总是首先寻求是否存在满足要求的已有产品或已有产品的变型,而不是首先考虑新设计一个满足当前要求的产品。已有产品的选择或变型就是产品变型设计的任务。变型设计就是寻求满足当前设计需求的已有产品,或对已有产品的动态模型的有效变型,形成新产品的过程。

前面已初步讨论了 CBR 技术在变型设计中的应用方法,并定义了三种变型设计。变型设计的实施主要是采用 RBR 和 CBR 技术,其中 CBR 是核心。前图 42.5-9 表示了实例推理过程和变型设计过程的关系。本节利用前面讨论的有关技术,论述三种变型设计的智能化实施方法。

### 4.1 产品定义变型设计

产品的变型设计可以发生在不同的设计阶段，因而存在不同的变型设计型式。产品定义变型设计是变型设计中的首选变型设计型式，因为它对企业的产品资源重用最多，对已有的产品动态模型破坏最小，因而设计的效率最高，对产品工艺和制造过程的影响也最小，它的优越性均是其他设计型式所无法比拟的。

产品定义变型设计是在实例检索获取的相似实例的基础之上进行的。这个检索过程从具有相似概念设计模型的产品实例（图 42.5-9 中 CAP 集合）检索开始，经过相似装配设计模型实例（图 42.5-9 中的 APS 集合）、相似产品定义模型实例的检索，最后形成了一个具有相似产品动态模型的实例集合（图 42.5-9 中的 DPS 集合）。从这个集合中提取最相似的实例（Ms-

DPS），并对它的产品定义模型实施必要的修改，直至满足用户的设计需求。

修改的方法是分解此产品相似实例。首先进行差异分析，以确定它和所要求的目标产品在何处不同，是装配体不同还是零件不同，然后检索不满足要求的装配体或零件的相似实例，若有可重用的装配体或零件实例，则经过组合后形成最终产品实例，否则修改之，如图 42.5-14 所示。

产品定义变型设计只是对组成产品的装配体内部零件及零件装配关系等的修改，不允许发生装配体的整体性质的改变。因为装配体的本质属性是在装配设计模型中定义的，若它的性质发生变化，则与其具有装配关系的其它装配体也会发生改变，导致产品的修改范围扩大；但若必须发生这种改变，则产品定义变型设计失败，开始产品装配变型设计过程。

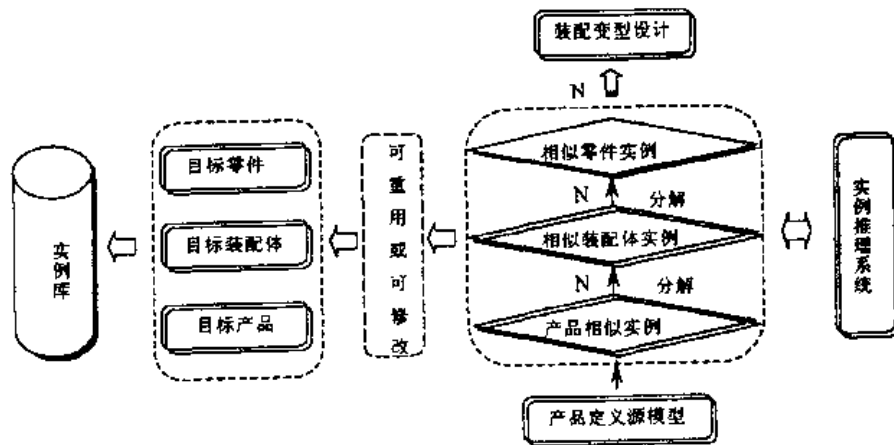


图 42.5-14 产品定义变型设计

### 4.2 产品装配变型设计

在产品变型设计中，如果一个目标产品无法通过

对相似产品实例的产品定义模型修改获取，则必须进行产品装配变型设计过程，如图 42.5-15 所示。

产品装配变型设计开始于从 APS 集合中（图

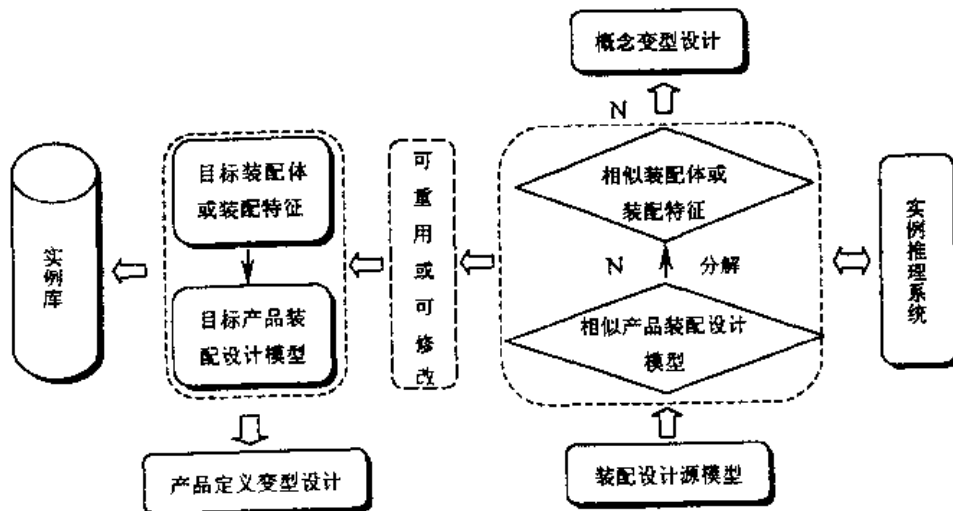


图 42.5-15 产品装配变型设计过程

42.5-9) 选出的具有相似装配设计模型的相似产品实例 Ms-APS。在经过差异度分析之后,确定需修改的对象,是装配结构层次关系视图中的装配体对象、装配特征,还是装配参数定义视图中的设计参数等。确定修改对象并经过修改之后,形成目标产品装配设计模型。设计的最终完成必须修改此实例的产品定义模型,形成目标产品定义模型。但确定与装配设计模型对应的产品定义模型则又是一个优选的过程,因为装配设计模型和产品定义模型之间是类和对象的关系,这实质上又进入了产品定义模型检索和变型设计过程。因此,装配变型设计的结果是新产品的装配设计模型和产品定义模型。

在装配变型设计中,不允许改变产品的功能、结构原理方案、机构体及它们的行为等属于产品概念设计模型的内容。若没有具有相似装配设计模型的产品实例,则产品的变型设计回溯到产品概念设计阶段,进入产品概念变型设计。

### 4.3 产品概念变型设计

变型设计一般首先进行产品定义变型设计及装配变型设计,在这两种变型设计均告失败的情况下,产品的变型设计就只能发生在概念设计阶段。发生在此阶段的变型设计虽然与其他两个设计阶段相比要复杂的多,但仍具有变型设计的意义。因为一般的变型毕竟不可能构成对现有产品大量信息的破坏,比如很可能是在原有产品功能的基础上增加一个功能。例如在原有液压支架功能的基础上增加一个护帮机构或铺底网机构。这种变型是无法在装配变型阶段或产品定义变型阶段完成的,因为产品的功能结构发生了变化。因此,也可以根据设计需求的情况,若发生了功能变化,则变

型设计必然要发生在概念设计阶段(这也说明在变型设计中,从产品定义变型,装配设计变型到概念设计变型,这样的顺序并不是一成不变的。完全可以根据具体变型设计的需求特点,直接确定变型设计阶段)。有时产品的功能结构虽然变化了,但并不是重大变化,只是打破局部的功能结构关系,而原有产品的大量信息仍可重用,这样的概念变型设计其速度和质量也就肯定优于全新设计。

概念变型设计始于实例检索获得的具有相似概念设计模型的产品实例 Ms-CPS (如图 42.5-9),实现过程如图 42.5-16 所示。

这个过程的重点是在确定了相似产品实例后,确定目标产品功能和相似产品实例功能的区别,目标产品功能向结构的映射要求和相似产品的功构映射视图的区别,从而定位修改的目标,实施修改操作,完成目标概念设计模型。

在概念变型设计中,可以增加产品的功能,改变实现任务功能的机构体及它们的行为,可以改变机构体之间的装配结构关系,选择不同的装配特征满足机构体装配关系设计要求。从这可以看出,产品概念变型设计是三种变型设计中设计变动允许范围最大的一个,在这个过程可以进行产品功能创新和实现功能的机构体创新,这也充分说明在变型设计中并不排斥产品的创新。产品的创新开发和产品的变型是相互依存的,创新之中有继承,而变型之中有发展。

目标概念设计模型确立之后,还必须确定产品的装配设计模型和产品定义模型,这就需要进行装配变型设计和产品定义变型设计两个变型设计过程。三种变型设计是循环往复进行的,产品的设计本来就是如此,这是设计的自身规律!

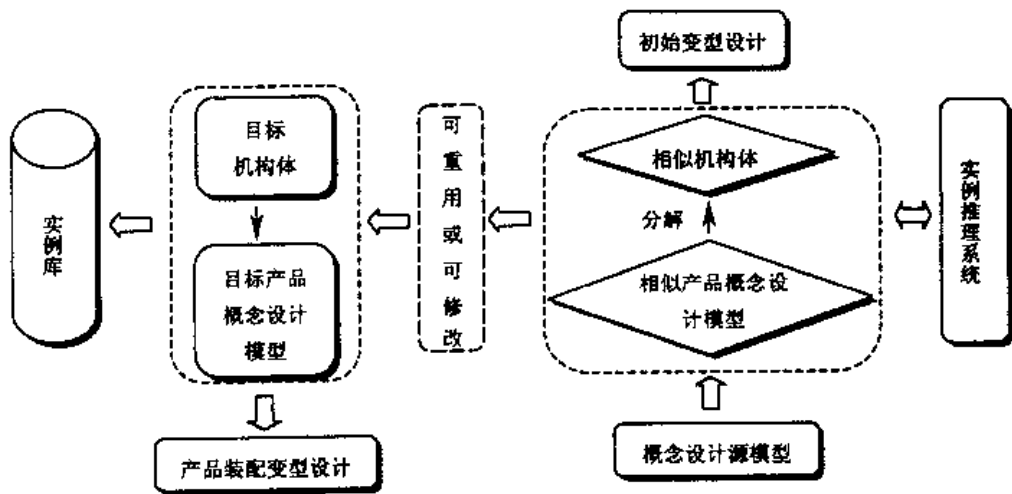


图 42.5-16 概念变型设计过程

## 第6章 实例研究

### 1 关系型产品模型的建模步骤

关系型建模是基于关系型 CAD 系统的产品建模方法。其核心思想是运用设计关系,将图(关系型变量化图形,或称白图、母图)、表(事物特性表)、码(类属模型的分属编码)三者有机结合起来,建立适合于整个产品生命周期的信息模型,以实现能快速响应市场需求的变型设计。

图 42-6-1 说明关系型产品建模的工作内容和操作步骤。具体如下:选择关系型 CAD 系统;进行零部件信息资源的标准化和规范化重组;借助于事物特性表技术,进行产品建模;交互式建立产品零部件母图。其中,事物特性表主要用来表征产品、部件、零件的功能特性、几何特性和制造特性等,并以表的形式反映出来。建表的方法和步骤是:通过分析特定对象类,去除不合理结构,抽取设计关系,以尽可能减少基本特性参数,而用导出特性(即表达式)来反映对象类的几何关系、拓扑关系和其他设计关系,最终形成一份包含图、表、码的设计文档(或称设计准则),以达到用固定的格式抽取并建立对象类的事物特性表。至于零部件母图,这种母图与传统 CAD 生成的图形相比有着本质的区别。首先,母图包含了大量的设计关系和设计约束,可记录设计师的设计意图和设计过程;其次,母图反映的是一类对象的信息集合,而不只是一个具体的对象实例;最后,有些母图还具有智能变型能力,使图上的标注尺寸、移位公差、尺寸公差和技术条件等,能随着特性表指针的变化,自动找到合适的方位和大小,

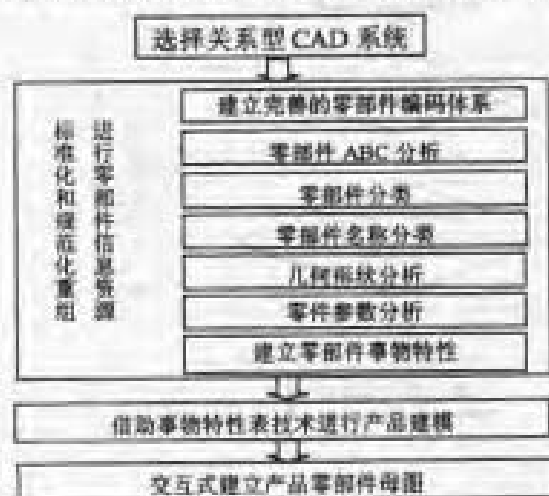


图 42-6-1 关系型产品建模步骤

以尽可能减少对变型所得零件工作图的再编辑和再修改。

下面列举几个实例,说明如何应用关系型产品建模技术实现产品的快速响应变型设计。

### 2 圆锥—圆柱行星齿轮减速机的变型设计

齿轮传动是机械行业中常用的变速装置,它通过齿轮啮合传递运动和动力。其中,圆锥—圆柱行星齿轮减速器采用两级齿轮传动的方式,输入级是锥齿轮传动,输出级是行星齿轮传动,特别适用于大功率、大速比又要求结构紧凑的情况。图 42-6-2 是传动副的示意图,其工作原理简述如下,由电动机输入的动力经锥齿轮轴 1 传到从动锥齿轮 2,经过一级输出轴 6,过摩擦离合器 7 输送给行星轮系中太阳(中心)轮 5,行星轮 3 和内齿轮(中心轮)2 的啮合传动,最后由行星轮 3 的系杆 4 驱动输出轴(图中未绘出),从而将高速级的输入机械能转换为低速级的输出动力。

根据前述的对象类类属模型的分属定义模型,此产品分类标识定为 30-2-1-3,它们依次表示产品类(自制,30),齿轮传动组(2),圆锥—圆柱行星齿轮减速器族(1),如图 42-6-2 所示的结构类(3)。按形状定义模型,此产品结构构成由三个部件组成:锥齿轮副部件(分类标识为 20-37-3-4)、行星传动传动副(20-37-7-1)和机架部件(20-45-2-3)。图 42-6-3 是它的事物特性定义模型。对于此类产品,它的基本技术参数为输入功率( $P$ )、输入转速( $n$ ),和总传动比  $i_{11}$ 。关于本例的算法定义,应包括基本参数计算(传动比、齿数分配)、几何计算(齿数、模数和其他几何参数)和强度计算。本文简单讨论传动比  $i_{11}$ ,

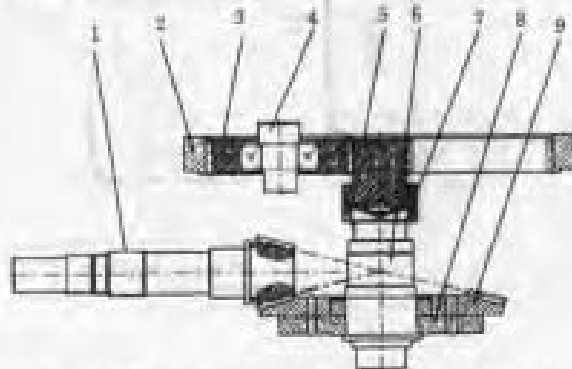


图 42-6-2 圆锥—圆柱行星齿轮传动结构示意图

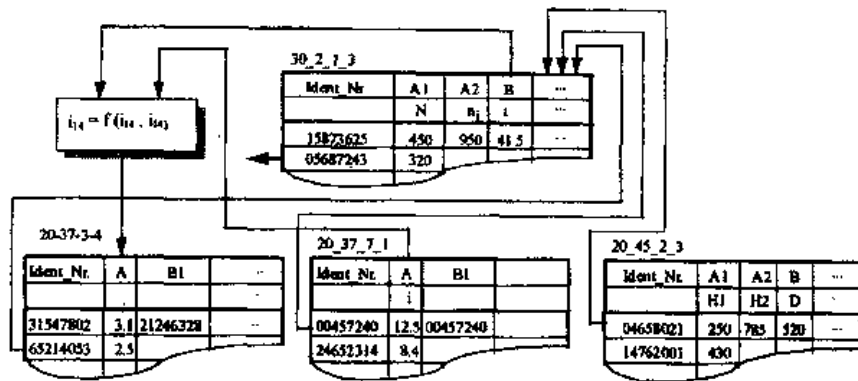


图 42.6-3 一个圆锥-圆柱行星齿轮减速机实例对象类的事物特性定义模型示例

的分配。

根据重组后的产品族类属模型，就能按用户的需求进行快速响应变型设计。其实质就是获取一个满足用户定单的实例对象。情况不外乎如下两种：一种是原对象类中已有合用的实例对象，这相当于已有产品的直接重用；另一种是重用对象类的属性定义通过参数重组产生新的实例对象。第一种情况自然简单。下面重点分析第二种情况的操作过程。对于一个选定的实例对象类，其实例对象的变化总是由输入功率  $P$ 、输入转速  $n_1$  和总传动比  $i_{14}$ （或输出转速  $n_2$ ）的取值组合的不同所致。从满足用户需求的  $P$ 、 $n_1$  和  $i_{11}$  出发，齿轮传动的基本技术特性  $z_1$ 、 $z_2$ 、 $z_3$ 、 $z_4$  以及模数  $m_1$ 、 $m_2$  等可以有多种组合，但由于它们的不同组合，会带来箱体尺寸的相应变化，势必引起新零件的大量出现。很显然，合理的变型设计结果应以圆锥齿轮传动中心距  $A_{19}$  和行星齿轮  $z_3$  中心距  $A$  为约束条件。同时，由于行星齿轮传动副中的固定齿圈是主机机架的一个组成部分，在整个产品中占有重要的地位，它的变化应受到严格的制约。因此，对用户需求的满足，首先选择预定义的优选  $A$  值，然后在保持  $A$  定值的条件下由所需的  $i_{19}$  ( $i_{19} = i_{14}/i_{34}$ ) 遍历实例对象类的实例对象，以找出一个能保证  $i_{14}$  位于给定波动范围内的  $(z_1, z_2)$  值。基于此，利用规范化、标准化的信息资源 ( $z_1, z_2$ )、( $z_3, z_4$ ) 系列值，就可以直接完成一个满足用户订单的变型设计。图 42.6-4 给出一个基于关系型产品模型的圆锥-圆柱行星齿轮减速机变型设计过程。其中，利用事物特性定义模型重用产品的属性定义或属性值的原理在图 42.6-3 有简略的表示。在图 42.6-3 中，当没有可直接重用的实例对象能满足用户需求时，可从 30\_2\_1\_3 中取出用户所需的  $i$  ( $i_{14}$ )，并从 20\_37\_7\_1 中取得一组 ( $z_3, z_4$ ) 和计算  $i_{34}$ ，再计算  $i_{19}$ 。又遍历 20\_37\_3\_4 中以找出能保持  $i_{19}$  位于波动范围的 ( $z_1, z_2$ ) 值。此时，可在 30\_2\_1\_3 中填入一行新的属性值记录，从而得到一个新的变型。

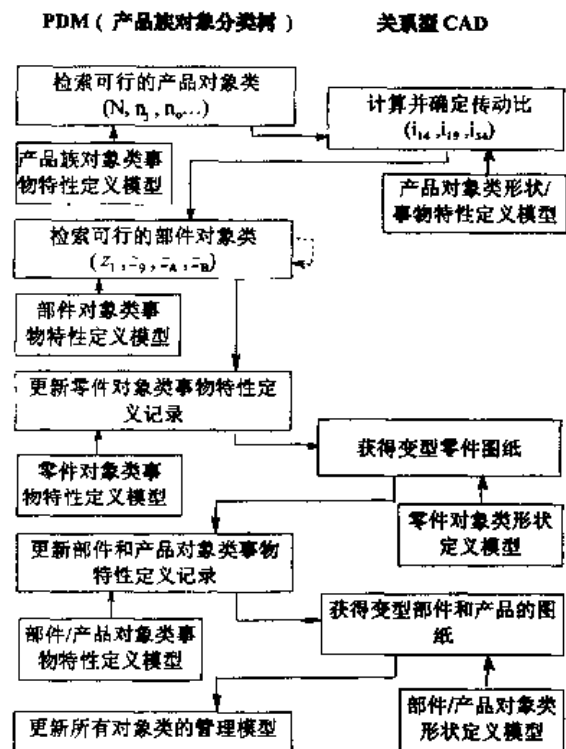


图 42.6-4 基于关系型产品模型的圆锥-圆柱行星齿轮减速器的变型设计过程

### 3 基于产品族的液压泵变型设计实例

产品族 (Product Family, PF) 设计实现产品供应可柔性，使基于族设计的可重用性达到最大效果，而不是一次只为一个产品设计，这为设计部门和销售部门特征化用户需求提供了基础，其方法是系统化的模块配置。整个产品族体系 (Product Family Architecture, PFA) 包括：基本构造块、配置规则、产品分类和经济评价。

基本构造块是构造变型产品的可重用和柔性的基础。通过重用设计能力、生产能力、加工能力、供应和制造资源，可获得设计和制造的经济性，也就是说，利

用产品模块化、可配置性、系列化,以及在范围和规模的折衷中获得经济性。基本构造块的确定原则是:用户当前和将来的需求;设计参数和生产过程的可重复性;易于配置。构造块的粒度等级由共用性和开发成本的平衡折衷来确定。

最终产品是由用户按配置规则装配构造块。配置的规则有:不同的构造块中的相容,例如构造块之间的干涉内容;用户需求和设计参数之间的映射关系。

产品分类可确定产品变型所需的基因产品。它反映了特别的产品战略和商业观点。基因产品体现了一类产品特点的公共性,一般产品就是在基因产品之上

附加其他构造块来组成。构造块的设计是通过特征和参数的选取来实现。通过产品族分类,用户可找到潜在地满足个人需求的构造块。

经济评价是基于公司的设计能力、制造能力、生产组织能力等评价产品族设计的成本和生产进度。可利用一些模糊规则来帮助用户折中决策。

PFA 是面向整个产品实现的过程。通过匹配需求 and 能力,实现用户价值最大化,同时通过增加规模来降低成本。PFA 的经济评价使得在产品族这一级优化,而不是在特定产品或产品子装配一级优化。产品族设计流程见图 42.6-5。

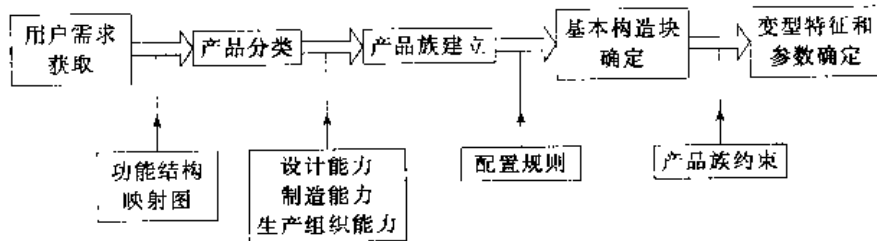


图 42.6-5 产品族设计流程

在液压泵产品开发中按照产品族的思想对经常修改的参数和特征建立参数表,把零件修改设计后的尺寸和特征填入参数表,生成实例零件,继承了原来的装配关系,然后替换装配体上的原零件,即可实现装配修改,并可进行干涉检验。原型系统的产品族设计可在

Visual Basic 上开发,装配分析可在 Pro/E 上作二次开发,产品变型特征和尺寸由产品设计专家系统、性能评价子系统及再设计专家系统来决定。变型特征文件经网络传至装配仿真子系统进行装配仿真。产品族设计及装配仿真流程见图 42.6-6。

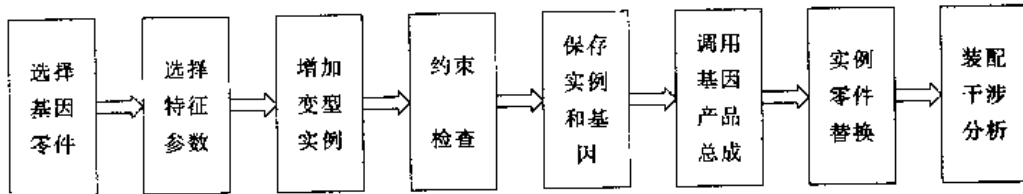


图 42.6-6 产品族设计及装配仿真流程

原型系统包括主菜单、产品结构树 BOM、以及 BOM 中当前激活的部件或零件图。在 BOM 中,可进行部件和零件的增加、删除和更名;如果激活的是部件,则显示该部件的装配图;如果激活的是零件,则显示该零件的基因零件图和尺寸标注,以及已有的变型零件名及尺寸值。主菜单有文档操作、尺寸关系、文档传输、窗口和帮助等。文档操作是对变型零件的实例进行操作,如选第一、选最后、向前一个、向后一个、移动子树、建立部件或零件、复制节点、删除节点以及保存等功能。尺寸关系中有该零件约束关系的建立、关系的读写、关系的增删和判断。尺寸判断是在当前的约束下检查变型零件尺寸关系的合理性。在选择“尺寸关系编辑器”命令前,必须选择一个零件。文档传输是把当前已生成的新变型零件通过文件远程传送协议(FTP,

File Transfer Protocol) 传输给产品设计服务器,进行虚拟装配分析。这样就能支持产品各部件的分布式异地设计和装配分析,在产品物理实现之前分析产品的装配特性。图像刷新的功能是双击图形,搜索并选择 acad.exe 文件,按“打开”按钮击活 AutoCAD。在 AutoCAD 中修改图形后,保存图形并将图形以 \*.bmp 文件输出至 bmp 子目录下,退出 AutoCAD,选择“图像刷新”命令,图形被刷新。选择“帮助”菜单下的“帮助主题”命令,弹出帮助菜单,选择“目录”,具体操作与 Windows 95 的帮助相同。原型系统中的工具栏同菜单功能。图 42.6-7 为产品族中变型零件设计时的变型参数输入。图 42.6-8 为产品族中的约束建立、检查和管理。



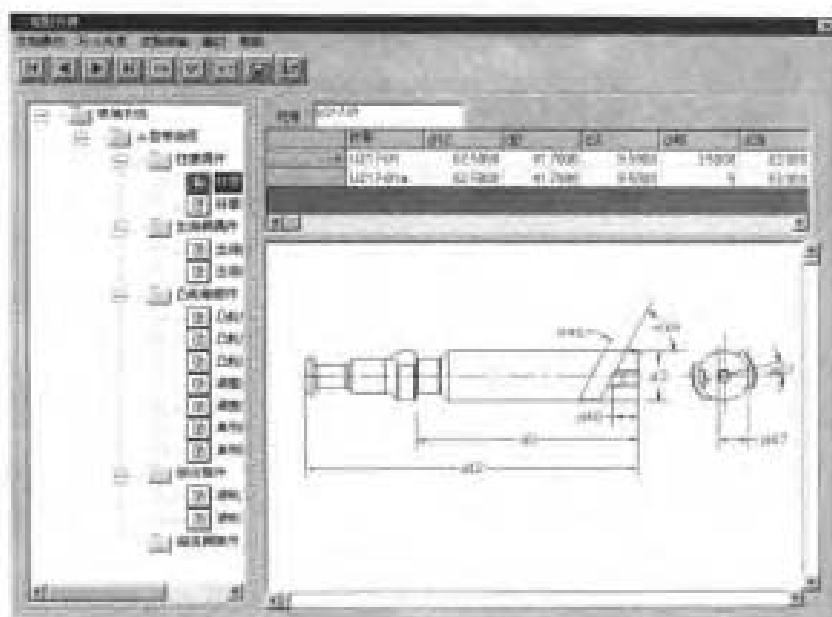


图 42-6-7 产品族中变型零件设计时的变型参数输入



图 42-6-8 产品族中的约束建立、检查和管理

# 第 43 篇 模块化设计

主 编 蒋寿伟  
编写人 蒋寿伟 刘溪涓

# 第 1 章 概 述

## 1 现代设计的特征

### 1.1 设计的概念

任何产品的问世均是设计的结果,大到航天设备,小到一个钮扣,无不是与设计密切相联的。从起点到某一终点应该如何走最快、最方便也是一种设计。因此设计是人类社会中一项非常重要的内容。但由于环境条件、接触产品的程度和范围不同,对设计的认识也是完全不同的,对设计本质的认识也有许多不同的观点,当然设计与社会及科学技术的发展更有分割不断的联系。在不同的时代对设计(包括形式、内容、方法、思想)的认识也将是完全不同的。从设计的发展历程来看,可分为三个阶段:

第一阶段:早期的很长一段时间内,人们把设计与制造看作是一回事,即边设计,边制造,设计、制造由一个人独立完成。古时的手工艺品、编织竹器等就是一例。这一阶段的设计的特点是设计与制造合一,设计是经验与构思的综合。

第二阶段:利用图样作为媒介,以传达设计的构思,1670年就出现了大海海船的设计图样。图样的出现使设计进入了思维交流的领域,制造也不局限于一个人,将有更多人投入同一产品的制造过程中。互换性原理的出现使这一阶段的内容越来越丰富,产品结构

亦趋于复杂化,满足的需求面也越来越广。这一阶段的特点是设计与制造的分。

第三阶段:计算机技术的发展和运用,特别是CAD/CAM/CAE技术更能把设计信息直接输出,以控制NC机床即可加工出所需要的零件,没有图样作为交流工具,产品的图形作为信息直接指导加工。这一阶段的特点是设计与制造的统一。

这三个阶段给人们的设计的认识产生了许多新的启示:

#### (1) 设计的有限时空性

任何超越当时的时空范围的设计是不可能的,也是无用的。设计的成果、形式、内容无一不带有时代的烙印。当然包括很多的设计思想。因此,考古学家对文物及发掘物的研究也是基于历史环境和历史条件。

如图 43.1-1 所示,以产品的质量为例:它是设计的一个重要内容,对质量认识的变迁也说明了不同时代对设计认识的差别。

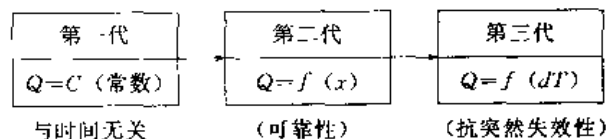


图 43.1-1 质量概念发展的模式

人们对质量态度的变化也就是一个晴雨表,反映了人们对质量概念的认识(如图 43.1-2 所示)。

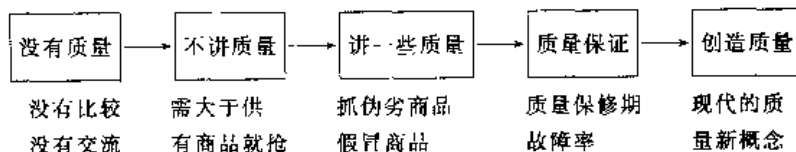


图 43.1-2 对质量概念的认识发展

根据著名的“七二一”规律,产品质量的 70% 取决于产品的设计及管理,产品质量的 20% 是加工制造,产品质量的 10% 是加工人员的素质。产品成本的 75%~80% 取决于产品的设计。

设计的有限时空性和社会性是设计的基础。

#### (2) 设计的需求性

设计就是为了满足某种需求,这就是设计的目的性。人们的需求从简单到复杂,从单一到多样的变化,体现了社会进步与科学技术的进步,人们需求变化的速率与社会进步的速率应是同步的。而需求的变化直接影响到产品的生存。从产品的生存期的变化可以看

到人们对需求变化的情况,产品的生存期越短,需求变化越快,需求产品的生存期越长变化也就越慢。过去我们的机电产品的生存期一般是 10~15 年,甚至更长,但现在机电产品的生存期 5~7 年,甚至更短。而轻工产品的生存期更短 1~2 年,玩具的平均生存期为三个月,彩电平均 2 年更换一代。生存期的缩短也就要求设计周期的缩短。

自行车从研究到定型用了 80 年。从 1818 年第一台脚踏式自行车到链条式自行车整整用了 170 年。

蒸汽机 19 世纪以前用了 100 年

- 电动机 1829~1886 年用了 57 年
- 电子管 1884~1915 年用了 31 年
- 汽车 1868~1895 年用了 27 年
- 雷达 1925~1940 年用了 15 年
- 电视机 1922~1934 年用了 12 年
- 原子反应堆 1939~1942 年用了 3 年
- 激光 只用了 1 年

(3) 设计的创造性

设计的创新是能不断地以新产品满足需求而不是用老的产品去满足新的需要。当然新产品的出现不是突然产生的,而是不断演变过程的积累。设计的创造性中包括了设计思想与设计方法。

设计的创造性是设计的本质。

(1) 设计的综合性

设计实际上是信息处理的过程,大量信息的积累、分析、研究和处理均是设计所需完成的基本工作。整个过程可看作是一个系统,按系统工程来分析设计过程中的每一环节,并予以协调,才能得到最优化的设计成果(如图 43.1-3 所示)。

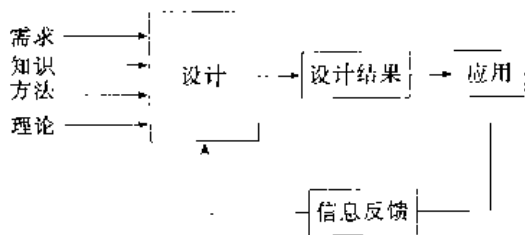


图 43.1-3 设计的系统工程

在信息处理中需用到许多相关的知识,它将综合应用各种方法、理论和知识来解决问题。

设计的综合性是设计的核心。

因此,设计应这样定义:在有限的时空范围内,在特定的物质条件下,人们为满足一定的需求而进行的一种创造性思维活动的实践过程。这种一定的需求将会转化为设计的依据——功能,以一定设计成果——产品来满足需求,这种需求就是要为完成某一特定动作的具体表现。为此,设计的结果是产品,产品总具有某一特定的功能,进行所谓的设计活动,也就是为了能生产出具有某一功能的产品的过程,这样的定义充分说明了其内涵。

设计是一个实践过程,不是某一环节,它是一个有时间要求的过程。在这一过程中的每一环节,因其目标不同,要求不同,对象不同,则其所用的方法也是不同的。这样,过程的长短就涉及到设计周期,对产品开发来说也就是开发周期。随着人们对产品需求变化的加速,产品的开发周期、设计周期也要大大缩短。

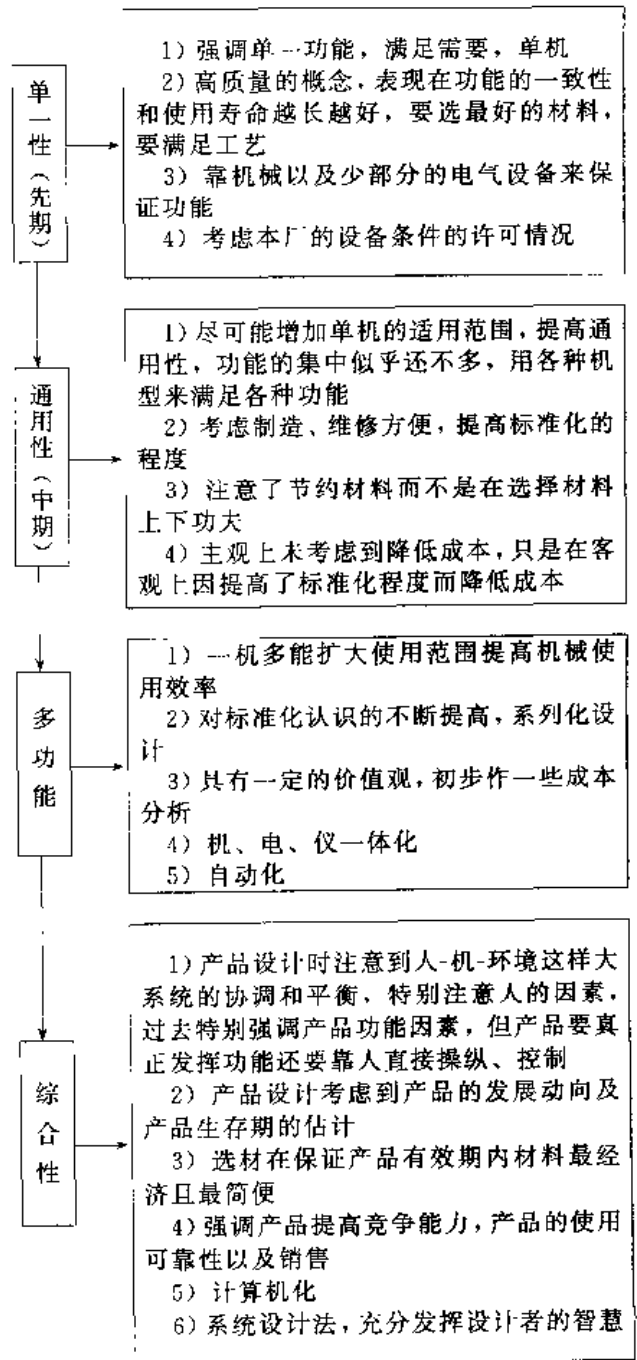
就以我们的机床行业来说:国内产品的开发周期

一般为 1~1.5 年,大型复杂的产品要 2~2.5 年,自动线需要 2~3 年;国外产品开发一般不超过 0.7 年,大型自动线 1.5 年

这种开发的差距不仅表现在设计方法方面,还有设计思想方面的。不同的设计思想有不同的设计方法,设计思想的变化是社会发展到某一阶段的体现,是对社会以及工作环境认识的深化。

1.2 设计思想的演变

设计思想演变的三段式



设计思想确定以后设计方法则是关键。

传统的设计方法是用数学、力学等基础理论知识进行静态计算,辅以设计者的经验和模型试验进行类比或经验设计。范围狭窄、动态信息不畅,产品开发迟缓,针对性差。而现代设计方法则是:

- 1) 用系统的观点研究设计的一般过程。
- 2) 提供具体设计问题的方法。

### 1.3 现代设计的特征

现代设计的五大特征:

- 1) 系统性 从系统角度上,对需求进行功能分析,使每一零部件具有显著的功能特征。
- 2) 交叉性 由于需求的多层次性及使用的多面性,则需要多学科知识的相互支持,学科的交叉使设计更具交叉性。
- 3) 针对性 按具体功能要求不同,必须进行针对性的设计这就体现在设计方法上,如对电子元器件则必须进行可靠性设计,而对低速、重载或高速的产品需进行摩擦、磨损设计。
- 4) 多样性 由于需求的多样性,现代设计就必须具有很大的柔性去满足多样性,以适应于时代发展的特征,一机多能的产品就是很好地体现。
- 5) 综合性 现代设计不再停留在静态的角度去设计,而是静态和动态的综合设计。

## 2 模块化设计的由来

模块一词来自于玩具积木。模块化作为一种现代设计方法,已经被人们所接受和应用,模块化又名模件化。

模块化的产生并不是偶然,而是有其历史演变过程的,这也是组合化发展到一定阶段的需要。模块化概念的完善和应用主要是从20世纪60年代开始,但模块化的思想则有着悠久的历史渊源。

自古以来,各种建筑物均是由一定形状和大小的砖和石料利用粘结材料经不同的组合而建成的。人类赖以自豪的金字塔,我国的万里长城也均是劳动人民利用简单的形状而构成形状、风格各异的建筑物,这就是组合化的雏型。组合化的出现还是较早的,除了建筑方面以外,人们早就开始把组合化的概念运用到日常生活中去,并扩展到工业生产的其他领域。1900年德国的一个家具公司设计出模块化“理想书架”,1920年前后欧洲一些工厂,运用模块化原理,用63个不同的齿轮组成60种不同的传动系统,用于丝杠、转塔车床等处,这些均是应用组合化原理的典范。19世纪20年代的包豪斯风格就是利用规则的几何体组合成各种产品,如勃兰特的茶壶、斯多菲的台灯、西门子的电话机

等。组合化也是经历了一个不断发展的过程。最早的组合,只是用于对基本几何形体的组合,如长方形的砖、圆柱形的梁等。这种组合的形状种类是有限的。从造型的角度来看也是很简单的,由于人们需求的不断增长,继而发展到具有一定复杂程度的形状,进行组合,如长矛、大刀、枪等,但它并不具有多功能的特性,而且发现仅是形状的组合尚不能满足多种的需要。为此,除了对自身形状复杂程度的提高,不但需形状组合,而且还需大小组合、方向组合、位置组合,例如组合夹具就是有这种特性。这种按功能要求提高自身形状的复杂程度,使各个基本件本身就具有一定的功能,再经组合后的功能也必然是多功能的。这样把组合化的概念又推进了一步,但上述的组合毕竟还是停留于单元的组合。一个单元所起作用毕竟是有限的,要是能再把单元进行组合,那么它就能具有更大的适应性。模块化也就应运而生。1967年英国曼彻斯特理工大学的Koeuigsberger教授对模块概念的阐述,为模块化应用的推广起了极大的推动作用。实际上在1963年日本的池田工厂就以大型机床为对象,使模块化实用化。与此同时,联邦德国也对此作了许多有益的探讨。模块化设计方法不但在建筑、机械、电子,甚至在教育系统中也大量应用。在高等教育中,由于专业的差异,课程开设的学时数要求也是不同的。为了进行有效的教学管理,把四年的课程按其共性和个性,利用模块化的原理,构成模块系统。该系统先按范围分别建立校级课程模块系统,然后再分解由各系按专业要求再建系级模块系统和专业级模块系统。政治思想教育模块、基础模块、基础技术模式、专业模式和实践模块系统等五大模块中各又有许多单元模块,这将分别由各系、各专业进行选择并予以组合,构成符合不同要求的教学计划和大纲。教育系统模块化的应用为进一步增大教育适应性,加强工程设计能力提供了有效的手段。

综上所述,模块化的产生主要基于如下的原因:

#### (1) 多样化时代的需要

随着科学技术和社会的发展,人们需求的层次越来越多,在广度上、深度上的变化都非常大。爱好、兴趣、功能差异的加大,单靠基本型的产品是不能满足需要,同时也不可能把所有的功能都集中在一个产品上,这样既不经济,也不合理。过去人们的一般观念受到了强有力的挑战。多样化时代的特点,可以用变、快、少来概括。

变:产品的多变以及不间断的更新换代是多样化时代的特点之一。单件手工加工向机械化、自动化生产的发展,完全适应了工业化时代的需求。它不但解决了产品的数量问题,更主要是能大大地提高产品的质量;但它却不能满足多样化时代的多变的需要,柔性化和

集成化则是解决中小批量生产和适应于多变的有效化手段,而模块化则是促进产品柔性化的巨大动力。

快:产品变化的速度要快,在一定时间内要有一定的新产品出现,这是衡量速度快慢的特性参数。这就要求开发品种设计的周期要短,生产周期也要短。目前,国外的工业先进国家的产品开发周期一般不超过1年,对大型自动线一般不超过2年。对塑料制品的产品,由构思、设计到投产的时间就更短,新加坡需10个月,美国需8个月,日本需5个月,而中国香港只要3个月,正是由于它们的快速、灵活使香港在国际市场的竞争中取得了明显优势。就以上海机床行业为例,要上一个品种的周期至少要1年,而模块化的快速、灵活正是模块化的最大特点。英国 Brunel Oni Kersity 等学校曾邀请了68位对计算机使用熟练程度不同的高级工程技术人员做了一个试验,对于复杂、一般、较简单的3个设计中的具体问题,按优化、模块化、线性化等三种设计思想用PASCAL语言分别编制程序,并在微机上求解,结果

在快速、灵活方面 模块化为37%  
线性化为17%  
优化为4%

在能得到最佳设计方面 优化为31%  
模块化为15%  
线性化为9%

由此可见,模块化在“快”字上是非常突出的。

少:批量小,而模块化由于可以使模块在最大程度上达到通用化,增大批量,以降低成本。

#### (2) 新生产格局的需要

我国一直沿用了20世纪50年代前苏联的以“大而全”或“小而全”为生产模式的生产格局,严重地影响了产品开发速度和生产周期,经济效益也就难以提高,我国的劳动生产率仅为国外的1/27。

以上海机床行业为例,其专业化程度低。工业发达国家和地区,其机床产品的零部件绝大部分是由专业厂提供,自制率仅占20%~30%,而上海机床行业的各厂,零部件绝大部分是本厂自制,一般自制率达到60%~70%。加之,制造工艺落后,人均劳动生产率低,设备落后(CNC机床仅占0.4%),经济效益必然是很低的。模块化将要求打破这种“大而全”或“小而全”的生产格局,要求以模块来组织生产。模块化工作的开

展也将有助于产业结构的调整、产品结构的调整,工厂、车间均可以模块来调整生产、工艺等,这样不但可以降低成本,同时可提高产品的质量。

所以,过去大批量、流水线的生产格局急需打破。

#### (3) 设计、生产周期的需要

由于产品生存期的缩短,更新换代速率的加大,也促使了模块化的出现。模块化的设计它不是一个产品一个产品进行的单独设计,而是一次设计就带出一批产品,再加上设计、生产周期的缩短,因而,通过不同的组合达到满足多变需求的模块化设计,就成为新型的先进方法也是必然的。

#### (4) 市场流通的需要

消费者的低价格、高质量的消费心理越来越强烈,因而,降低成本也就成为各制造厂关注的焦点。模块化设计在这方面将有着独特的优点。联邦德国马克公司用模块化设计制造单梁吊,其设计费用降低了12%,生产成本也降低了45%。因此,模块化体现了极大的优越性。

#### (5) 功能的需要

集成化和微型化的影响,要求产品不仅能做到零、部件互换,更要做到不同功能的组件也能互换,以使整台产品既具有多功能,又不扩大产品的体积和重量。在这方面,电器的元器件、计算机等均走在机电、造船的前面。集成电路、超大规模集成电路的出现,实际上是模块化的体现。现代电视机、高精度的测试仪器无一不是采用单块接插(具有不同功能的)组合而成。微机中留有各插板位置,当加入附件时能使计算机的功能得以增加,这些均是模块化的直接作用。

系统工程的出现,从理论上为模块化奠定了基础。每个系统均是可分的,可分为若干个子系统,而各子系统又可分为若干单元。若按功能的要求,把总功能分为若干子功能,并与子系统相对应,那么在组成系统时,经过不同的组合,可得多种不同的功能要求。这种子系统我们可以把它作为模块,那么总系统就是模块系统。模块系统只有在系统工程理论的指导下才能得发展。

综上所述,尽管模块化思想早就在生产实践中得到应用,同时在不同的领域也在自觉地应用它,但模块化真正得到确认、发展、成熟及应用,还是在上述各个条件的影响和指导下进行的。

## 第 2 章 模块化设计的基础——功能分析

### 1 功能分析的概念

设计的目的是要制造出功能全、性能好、成本低、结构先进的产品。产品的第一目标应是功能,功能是产品的根本,其次才是外观的造型。没有达到预定功能的产品只是艺术品,只能是人们观赏的对象。而具有最佳功能,外形简陋而又难看的产品,则会由于操作者容易产生思想疲劳而影响生产效率的发挥,对产品而言也会影响或失去其竞争力。功能与造型在某种程度上会产生矛盾,在这种情况下权重利弊,则应是以保证功能为主,并确保具有一定美的外形。

功能是产品所具有的特定用途与各种特性的综合,也是产品能完成任务的抽象描述。功能的具体化则是靠产品的技术系统予以体现和实现的。功能分析则是技术系统设计的主要依据,只有用功能的观点来认识产品的本质才能摆脱原有结构的模式而创造出更多、更新的方案,并取得最佳的结果。

产品的功能是由其技术系统来体现的。一般来说,

技术系统整体的功能要求是该系统应具有的总体功能

例如:车床用于车削。

汽车用于物料运输。

所谓功能,就是完成物质、能量、信息、生物四种流的作用。

物质流:指材料的形状、大小的变化。

这里所指材料:是泛指如饲料、皮革、食品、气体、液体。

能量流:指能量的变化。

电能转化为机械能,电能转化为化学能,机械能转化为热能,以及光能、核能、生物能的转化。

信息流:数据、控制脉冲的显示。

生物流:是指城市建设、环境保护、建筑。

四种流的不同也就区分了机、器、仪,见表 43.2-1。

实际上,技术系统的功能是指其把输入量转变为输出量的能力(如图 43.2-1 所示)。

总系统——总功能。

表 43.2-1 机、器、仪的划分

主要传递	机		器		仪器
	能量及物料				信号
对象	机(系统之间有相对运动)		器(系统之间无相对运动)		
主要性能	强度、刚度、寿命、磨损		静强度、刚度		精度、可靠性
实例	电动机 燃气轮机	机床 交通工具 运输机	锅炉 冰箱	高炉 管道系统	计算机 打字机 测量仪器
	主要转换能量 ——动力机	主要转换物料 ——工作机	主要转换能量	主要转换物料	

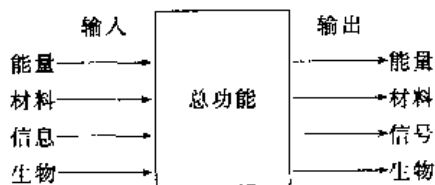


图 43.2-1 功能示意

部分系统——部分功能——模块的功能

以自行车作为一个系统(如图 43.2-2 所示),进行分析:

自行车的边界条件是骑车环境,输入是能量——骑车者的脚蹬力,输出也是能量——高速度的承载力。

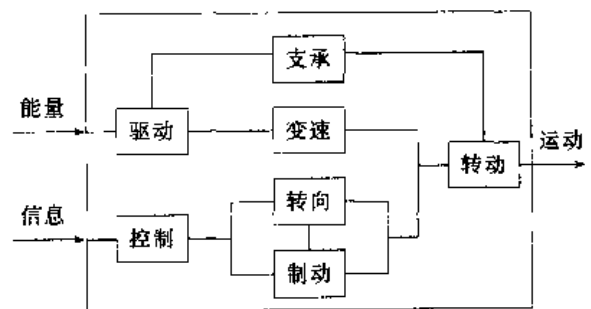


图 43.2-2 自行车的技术系统

由输入和输出之间的差异及关系可以分析出自行车系统的总功能是能量的转换,骑车者借助自行车系统用

自身力可得到一定运动速度下的承载能力。

## 2 功能的分类

功能的分类可有几种方法，按功能的性质及其重要性可有如下的分类：

### (1) 按功能的重要程度分类

基本功能：完成总功能所必需的分功能。

辅助功能：为实现基本功能起保证作用的分功能。

在基本功能中要注意关键功能；对关键功能将采用特殊的技术手段（Know How）使之具有新的水平。

### (2) 按用途分类

驱动功能、传动功能、控制功能、工作功能。

### (3) 按功能的性质分类

使用功能：设计对象的适用性、可靠性、安全性、维修性。

美学功能：外观功能、艺术功能、产品的造型色彩、图案和装潢。

### (4) 按用户的要求分类

必要功能：完成总功能不可缺少的分功能。

非必要功能：多余的、过剩的功能。

主体功能：产品的目标功能。

附加功能：与产品的目标功能无多大关联的其他功能。如电扇装有时间显示器等。

### (5) 按其作用分类

通用功能：概括了社会的一切事物而抽象出来的具有通用性的功能。通用功能如转换、离合、联接、导向、存储等。

专用功能：除了通用功能以外的功能。

### (6) 按其原理分类

物理功能：能量的转换。

逻辑功能：可编程序控制器。

数学功能：差动轮系、核子秤、加、减、乘、除及开方等。

由于功能的可分性，总功能可再分为功能元，对功能元来说，在某些场合下这种按原理的分类可能更为有利。

例如物理功能元：反映系统中能量、物料及信号变化的物理基本作用。法国 R. koller 教授总结出技术系统中物理功能的 12 对典型基本动作。

## 3 功能分析的涵义

功能分析是技术系统设计的前提，也是产品符合原定需求的基础。功能分析的错误将导致设计出来的产品大量积压和报废。

功能分析是指：

1) 功能分解。按预定的总功能分解为若干个分功

能及功能元

2) 功能求解。按功能分解图以选择其功能载体和解法原理。

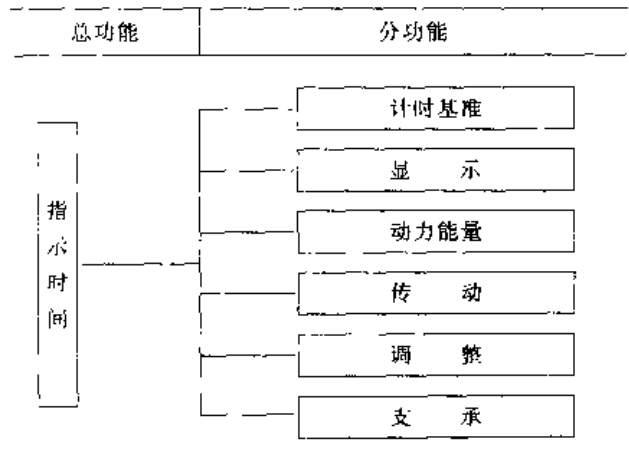


图 43.2-3 钟表的功能分解图

图 43.2-3 是钟表的功能分解图，为满足上述的分功能可有多数的原理解，而每一种原理解又可以用多种的功能载体予以保证，这些可以通过设计目录进行选择、优化，组合后可得到各种系列的钟表产品。

表 43.2-2 所示为钟表产品而编制的设计目录。

表 43.2-2 设计目录

分功能	解法原理	功能载体
计时基准	机械弹性摆 晶体振荡电磁摆	手表用弹性摆 钟用长摆 石英晶体振荡器
时间指示	视觉显示 听觉报时	刻度、指针 石英数字管 声音发生器
动力、能量	弹性变形产生能量 化学能→电能	游丝弹簧 微型电池
传动	机械传动	齿轮
调整	手动	旋钮或直接转动指针
支承	壳体	不同材料 不同的要求 使用的环境要求

在设计目录中，按其分功能不同的解法原理和不同的功能载体进行组合则产生了座钟、台钟、闹钟、手表、挂表等系列钟表产品。

由图 43.2-3、表 43.2-2 可见，功能分析对产品设计的巨大作用，由此变型产品开发有了方向和目标。对手表厂来说，其产品的总功能是不变的，分功能也是不变的，只要改变解法原理和功能载体就可有一系列的变型产品出现。

在现代产品设计中，一般均是从功能分析入手，以求得最佳原理方案。功能分析是方案设计中一个首要



的重要步骤，其步骤与方法如图 43.2-4 所示。

在分析系统总功能时常采用黑箱法。

所谓黑箱法：对一些复杂的系统和问题，人们很难一下认识模拟和控制，犹如一个不透明、不知其内部结构的“黑箱”。对其通过外部观测，分析黑箱与周围环境的信息联系，了解其功能、特性，以进一步寻求了解其内部机理及结构。如图 43.2-5 所示。

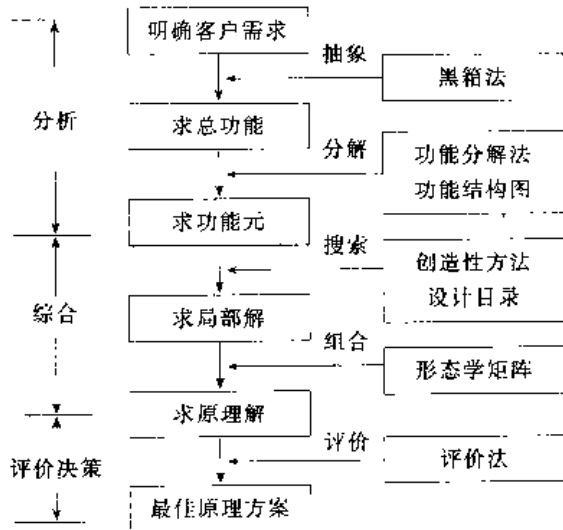


图 43.2-4 功能分析步骤

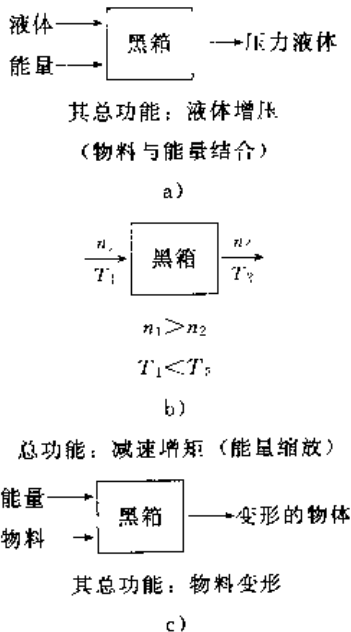


图 43.2-5 黑箱法

在此基础上选择理想的方案，则 Black box 就变成 glass box。

总功能的分解，可利用功能分析图来进行，如图 43.2-6 所示。

$F$ ：总功能； $F_1, F_2$ ：分功能； $F_{11}, \dots, F_{23}$ ：二级功能； $F_{121}, \dots, F_{212}$ ：三级功能； $F_{2121}, \dots, F_{2124}$ ：功能元。

功能元是功能分析中的最小基本单元，它是具体对应于原理解的不能再分解的功能，经它们的组合可得上一级分功能的解，直至取得总系统的原理解。

从理论上而言，功能元是功能的最小基本单元，即不再可分。划分功能元是为了找到相应的原理解和功能载体。如果已经能找到相应的，而且是满意的原理解及功能载体，而功能尚还可再分小的话，此时不必再继续再分，可认为它已是基本单元。所以功能元的复杂程度也是不同的，这要与具体功能载体的成熟程度而定。

功能载体即为实现对应功能的实体结构，如齿轮可以实现动力和运动的传递功能，则其就是功能载体。

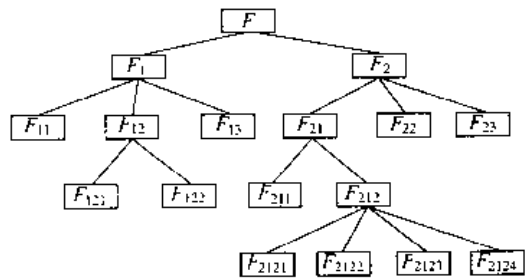


图 43.2-6 功能分析图

在总功能相同的情况下，通过功能分解去寻求功能元，其分解结果不同，则相应的功能载体的解决方案也是不同的，这就存在相互比较和选择。

如车削加工中工件的定位与夹紧，若不分开用三爪自定心卡盘，若分开的则用四爪单动卡盘。

功能是客户需求的反映，是技术系统的目标，也是设计的依据。在产品设计中，总功能确定以后功能分解就成为重要的问题。功能分解是方案设计的基础，按不同的分功能或功能元选择不同的解法原理及功能载体。

图 43.2-7 所示的是功能分解的流程。

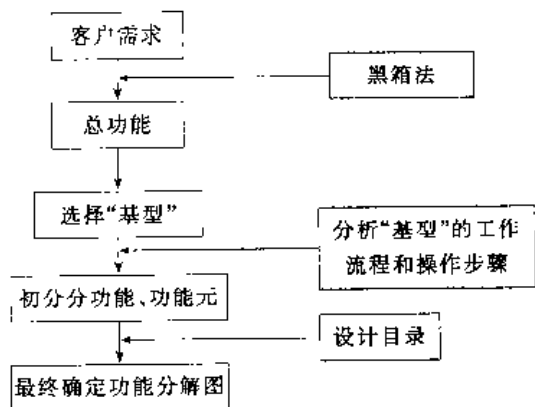


图 43.2-7 功能分解的流程

功能分解的步骤：

- 1) 用黑箱法求解总功能，对引进的产品，则可用反求技术，来求得总功能。
- 2) 选择与该总功能的相似物——“基型”。

- 3) 分析“基型”的工作流程和操作步骤。
- 4) 按流程和步骤确定分功能或功能元。
- 5) 按分功能和功能元初选解法原理和功能载体。
- 6) 归并、整理分功能和功能元，并最终确定。

### 4 功能结构图

#### 4.1 功能结构

技术系统的功能也是逐级对应的。与总系统相对应的是总功能，与部分系统或子系统相对应的是子功能。这部分功能也就相当于模块的功能。

功能之间按照确定的关系相结合，才能产生总功能。把功能及其之间的关系总和称为功能结构。

功能结构是功能之间关系的集中体现，它们之间不但有顺序，逻辑关系，也有最佳关系的优化、选择。功能结构一般有四种基本型式，如图 43.2-8 所示。

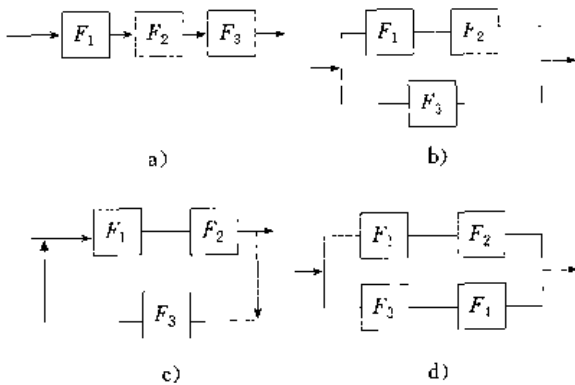


图 43.2-8 功能结构关系的三种基本型式

- a) 链式结构 (串联)
- b) 并列结构 (并联)
- c) 循环结构 (回路)
- d) 混联结构 (混联)

图 43.2-8a 所示是链式结构，是各分功能按顺序相继作用具有一定的逻辑关系。图 43.2-8b 所示是并列结构，是各分功能并列作用（同时作用），否则是不能产生输入和输出的变化。图 43.2-8c 所示是循环结构回路，具有反馈信息的作用。而图 43.2-8d 所示则是一种混联结构，即两个串联结构之间再用并联的形式组成一个较大的功能结构系统。

建立某一产品的功能结构必须注意：

(1) 在功能结构中的功能的属性

在技术系统中，输入和输出对象如材料、能量、信号等均是物理量，若功能结构中输入和输出量属于物理关系的改变，则以建立物理关系的功能结构为好，如车床的功能结构即是如此。但若功能之间的关系是属于逻辑关系的话，则以建立逻辑关系的功能结构为好，这种逻辑关系即是“与”、“或”及“非”的三种基本逻辑关系，如汽车的车窗的门锁。其次尚有数学关系的功能结构，即指加、减、乘、除和乘方、开方、积分和微分等。

(2) 功能结构的优化组合

不同的功能结构将会构成不同的技术系统，选择最佳的功能结构是技术系统发挥功能作用的重要保证，这就需要对功能结构从功能的发挥及其复杂程度上进行总体评价。

(3) 功能结构的建立是经反复不断的比较修整而形成最终的功能结构

在最初建立功能结构时，要按“基型”或相似的参考物的功能结构为参考，并以此经反复比较组合才建立最终的功能结构

#### 4.2 功能结构图的建立

功能结构图是用于描述系统的输入及输出，实现其总功能时其中分功能和功能元及其之间的顺序关系的图形。它的建立能使技术系统具体化，是整个技术系统虚拟的前提和依据，同时也要体现这些分功能和功能元之间是物理关系、数学关系还是逻辑关系。功能结构图也是为技术设计中机构的连接结构图提供依据，连接结构图则是把各分功能和功能元的功能载体之间的连接形式进行描述的图形。

对功能结构图应具有如下的要求：

- 体现分功能或功能元之间的顺序关系，这是与体现分功能和功能元与总功能之间的层次关系的功能分解图的重要区别。

- 体现分功能和功能元之间的属性即物理关系、数学关系还是逻辑关系。

- 体现出按此功能结构图能否满足总功能的要求。

- 体现出分功能和功能元之间的主次关系。哪些是主要功能、哪些是次要功能，或者哪些是基本功能、哪些是辅助功能，而哪些又是附加功能。当某些产品添加一些有实用意义的附加功能时，产品就能得到高附加值。

以台式钻床为例确定其总功能，画出功能分解图和功能结构图。

- 1) 以黑箱法求出其总功能，如图 43.2-9 所示。

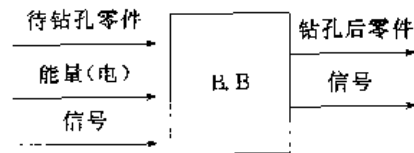


图 43.2-9 总功能：零件钻孔

- 2) 画出功能分解图，如图 43.2-10 所示。

- 3) 建立功能结构图，如图 43.2-11 所示。

功能分解后想求出系统最佳原理方案解，就必须对各种原理方案进行评价。评价方法较多，一般最常用

的是加权评分法（有效值法），其工作步骤如图 43.2-12 所示。

加权系数是评价目标重要程度的变化系数，目标

越重要，加权系数越大： $g_1 < 1$   $\sum_{i=1}^n g_i = 1$

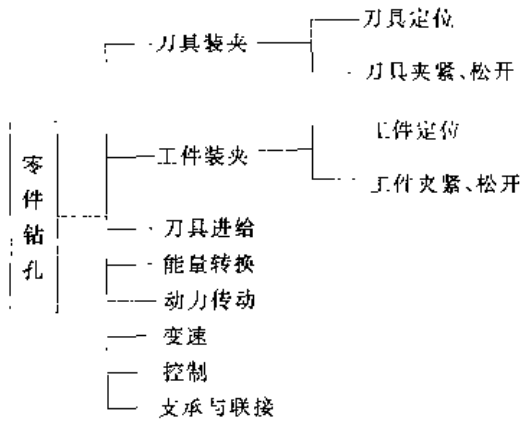


图 43.2-10 台式钻床功能分解图

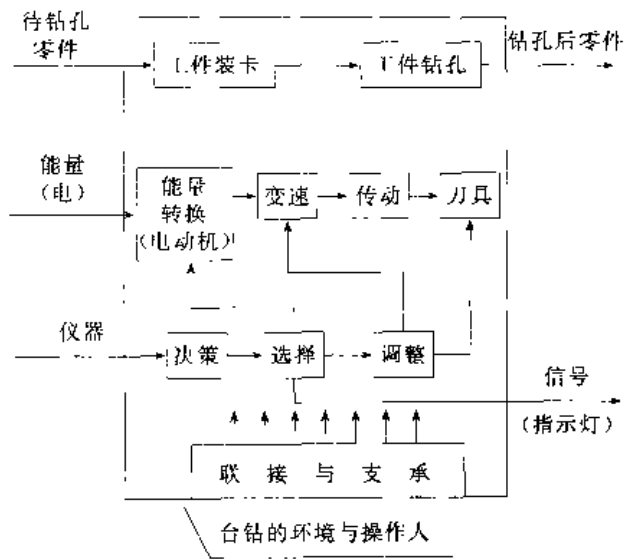


图 43.2-11 台式钻床的功能结构图

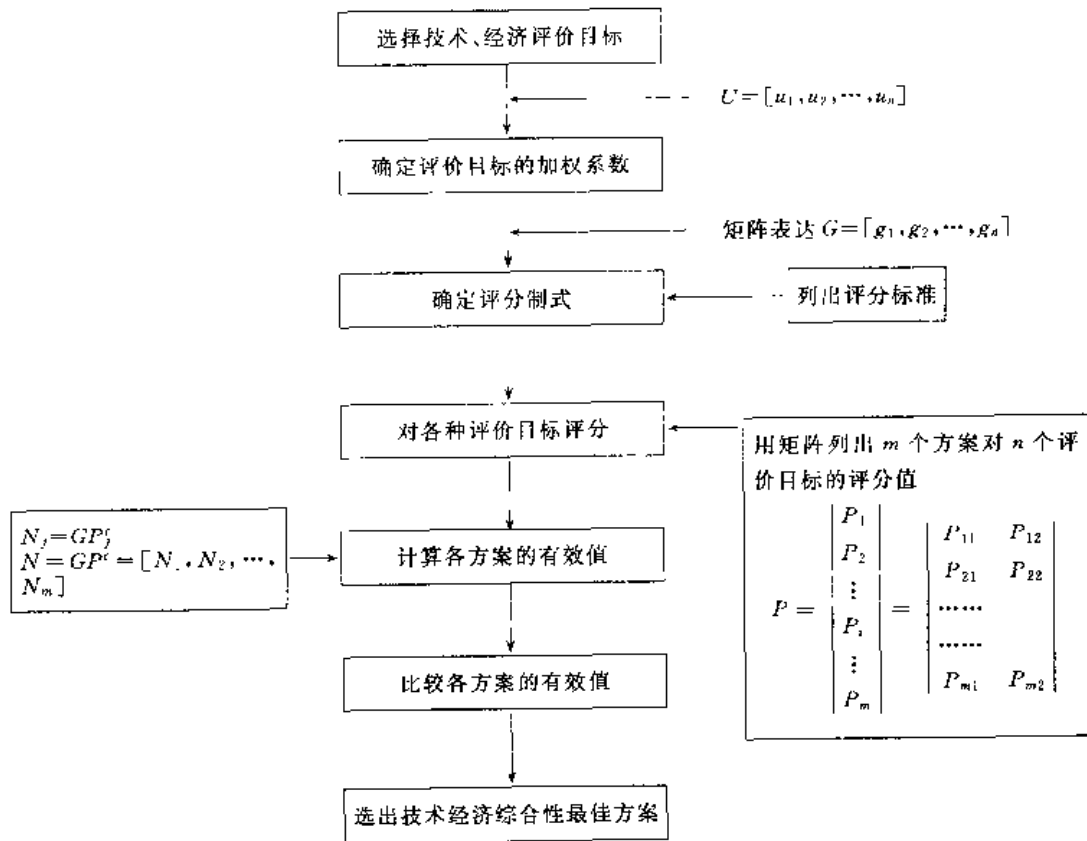


图 43.2-12 加权评分法的工作步骤

## 第 3 章 模块化设计基础

### 1 模块与模块化

#### 1.1 什么是模块化 (Modularization)

模块化又称模件化,也有把它称为积木化。如机械工程词典中把模块化程序(modularized program)也称为积木式程序。“模块化程序,它基本上是将一个大型程序从结构上分解成许多相对独立的模块程序。而每一模块程序具有特定功能,它既有信息入口,又有信息出口。这样就可能使大型程序结构清晰,调试时可以分块进行。……不同模块程序的组合就可构成不同的大型程序……。”

模块化的定义有很多,如表 43.3-1 所示,但按其概念应具有如下的定义:由若干具有不同用途(或性能)并可互换的模块,经不同的组合,以满足不同需要的这种方法称之为模块化。由此可见,模块化应具有四个基本含义:

(1) 必须具有一定数量的模块

所谓模块:是指一组具有同一功能的联接要素,但有不同用途(或性能)和结构,能够互换的各个单元。模块是组成模块化系统的基本单元,模块可以由一个零件组成的,也可能是一个组件或部件。一个产品应分成几个模块及模块中应包括哪些零件等问题均是在模块化中很重要的问题。要成为模块必须具备四个条件:

1) 具有独立的功能。每一模块都具有自己特定的功能,该功能是总功能中的一个组成部分,这种功能将可以单独进行调试。

2) 具有联接要素。每一模块与其他模块组合后体现总功能要求,但这种组合不是简单的叠加,而是通过一定的联接形式来进行组合,为使每一模块的联接形式通用,这种联接形式应是标准化的。

3) 具有互换性。各模块之间应是可以互换的,这样才能便于各模块之间的组合,以得到不同功能的要求。联接要素的互换性是模块互换性的组成部分,而更主要的是不同功能模块组合后仍可满足原定的需要。例如车床的模块系统,其他模块不动,把尾架顶针模块用带有钻头的钻削模块代替,那么,车床增加了钻削的功能并保持车床的原有功能。

4) 是一组有不同用途的,不同结构的基本单元,这里需要特别强调的是——组,而不是单一的,因为只有

表 43.3-1 模块的定义汇总

依 据	定 义
Collins Dictionary Of the English Language	标准的整装单元,例如电子元件的组(合)件或计算机硬件,它可与其他单元一起使用。从宇航学来说,组成宇宙飞船或运载火箭的几个整装可分单元,其中每一个都有一个或几个特定的任务
Webster's New Collegiate Dictionary	1. 一系列用在一起的标准化的单元中的任一个 2. 通常由(许多)电子元件整装的功能组件,供与其他这类组件一起使用
英汉微型计算机辞典	按一定标准设计制造的硬部件或软件,它在系统中易于代换和使用系统容易扩充
机电工业辞典	是指一组具有同一功能的结合要素(指联接部分的形状、尺寸、配合等),但有不同用途(或性能)和结构的,能够互换的各个单元(零件、组件、部件或系统等),或是能增加机械功能的单元。而单元是指能发挥特定功能的完整系统,如设备或建筑物
英汉技术教育辞典	模块化 (Modular) 模块化, 积木化 (Modularization) 微型组件
日汉科技词典	模件、模数、微型组件
Webster's New WORLD Dictionary	1. A Detachable unit with a special function, as in a spacecraft. 带有某一特定功能的可分单元,如宇宙飞船 2. Any of a set of units to be variously fitted together 各种相互配合的单元中任何一个部分

是一组不同用途、不同结构的基本单元,才可以有选择地经不同的组合后达到不同的功能要求以适应不同的需要。若是单一的就没有选择余地,那么组合后的功能也就不可能是有所区别的。而这一组的基本单元就是模块,这些模块不应是大小上的区别,而应是用途和结

构上的区别。

上述四个条件是区别模块和非模块的基本条件,这里必须指出:构成模块的基件之间的连接形式并不是区别模块与非模块的条件。模块可以是可拆的,也可以是不可拆的。同时也不要为了运输的方便,而分解的组件看作是模块。例如,车床的顶尖、铣刀、砂轮等等。

#### (2) 应用系统组合原理

模块系统的多功能不像其他产品在不调换附件的情况下直接体现。模块系统的多功能是通过不同模块的组合体现的,因此模块系统的产品必须应用系统组合论原理,按需组合。这种组合的关键是必须保证各模块联接要素的一致性。

#### (3) 最终要获得能基本满足各种不同功能的需要

模块系统的功能是由各模块组合后体现的,故不同的功能模块经组合后体现出不同的功能。总的来说,模块系统是多功能的。这种不同的功能,可以体现在具体性能的不同。如 Shanghai200 模块化仪表车床,该系列共设计了 99 个模块,若选用其中 54 个基本模块,至少组成 202 种产品。CO320 仪表六角车床,CO420 仪表卡盘车床,CO520 仪表精整车床,仅是上述组合后的三个仪表车床。

通过调换模块可改变加工的方法(这也是功能要求)。如高精度外圆磨床经模块化设计后,它可以自动纵磨,也可自动切入磨;既可磨削外圆,又可磨削内圆;可定程磨削,也可用量仪控制自动磨削;可磨削圆柱表面,也可磨削圆锥表面。

通过调换模块尚可满足各种主参数的不同需要。如高精度外圆磨床其磨削长度 500mm 及磨削直径  $\phi 200\text{mm}$  作为基型。通过模块化设计后,磨削长度可为 350mm、500mm、750mm、1000mm,而直径为  $\phi 200\text{mm}$ 、 $\phi 320\text{mm}$  等。

#### (4) 模块化的可分性

模块化系统可分为若干个模块,而每个模块均有其独立性,对用计算机进行模块库的建立和管理是相当有利。同时,也给维修带来了方便。

## 1.2 模块化的作用

模块化的作用主要是:

1) 由原单一功能的产品变成多功能产品。一般的卧式铣床通过模块组合可成为多功能机床,如卧铣、立铣、精密钻等。

2) 便于应用成组技术(GT)及CAD。

3) 增加批量,降低成本,大大缩短设计、生产与制造周期。

机床厂过去在单台产品设计时,三人一年半设计

一台,而采用模块化设计后平均每人每年可完成 1.5 台新产品设计,大大缩短了设计周期。

4) 便于计算机管理。通过建立模块化系统,把该模块的形状、大小等储存于计算机中,用户可通过显示器按需组合,直到满意,最后通过计算机绘出组装图,列出模块明细表等,然后按清单到仓库把有关模块选出并组合成产品,通过计算机管理还能对模块的组合给出最佳的组合方案。

5) 增强了国际市场的竞争能力。由于模块化能大大增强制造厂对市场的应变能力,因此也同样增强了它的竞争能力。

6) 便于产品更新换代,发展变型产品。不断地运用新技术设计出高性能模块,以代替原有模块,并组合出具有新功能的模块系统。

## 2 模块的划分原则

按模块定义可知:模块具有独立性、功能性、互换性和成组性,也就是每一模块将是与它的功能相对应的,而模块所具有的功能也必须与总功能相匹配,对于总功能中所不具有的分功能或功能,与之相应的模块也是不存在的,也就是在技术系统中是不需要的。由于功能的可分性,下一级功能是上一级功能的保证,那么与之相应的模块也就存在着上一级与下一级或低一级与高一级的模块之分。下一级模块是上一级模块的重要组成部分,这样的层次结构将会有利于发挥模块化的作用。

在实际模块的划分中往往不能以零件和部件作为区分界限,零件不一定能成为模块,这要由其体现的功能是否是所需的功能来决定的。在模块化设计中,模块的划分是一个非常重要的问题,模块划分的好坏直接影响到整个模块化系统的复杂程度,结合与分离的方便性,装配质量的控制以及组合后所产生的产品类型数量,同时,也直接影响到经济效益。

在一般的模块化设计中,对模块的划分将遵循如下的基本原则:

1) 模块应具有自身的独立性。它不应依附在其他模块上而应完整地独立存在,以完成规定的功能。这样的模块易于组合成各种变型产品,如磨床主轴轴系设计成整体套筒结构即是一例。

2) 要用有限数目的模块构成大量的不同组合的产品。这与块规的组合具有相同的性质。如 83 块一组的量块可用不多于 4 块量块的组合来复现 0.5~340mm 的任意长度。但若采用砵码组的数系来构成长度间隔,那为复现 188.88mm 则需要 13 块量规。模块的数目与模块的划分有着密切联系。

模块的数目将随模块自身的情况而定,如因基础

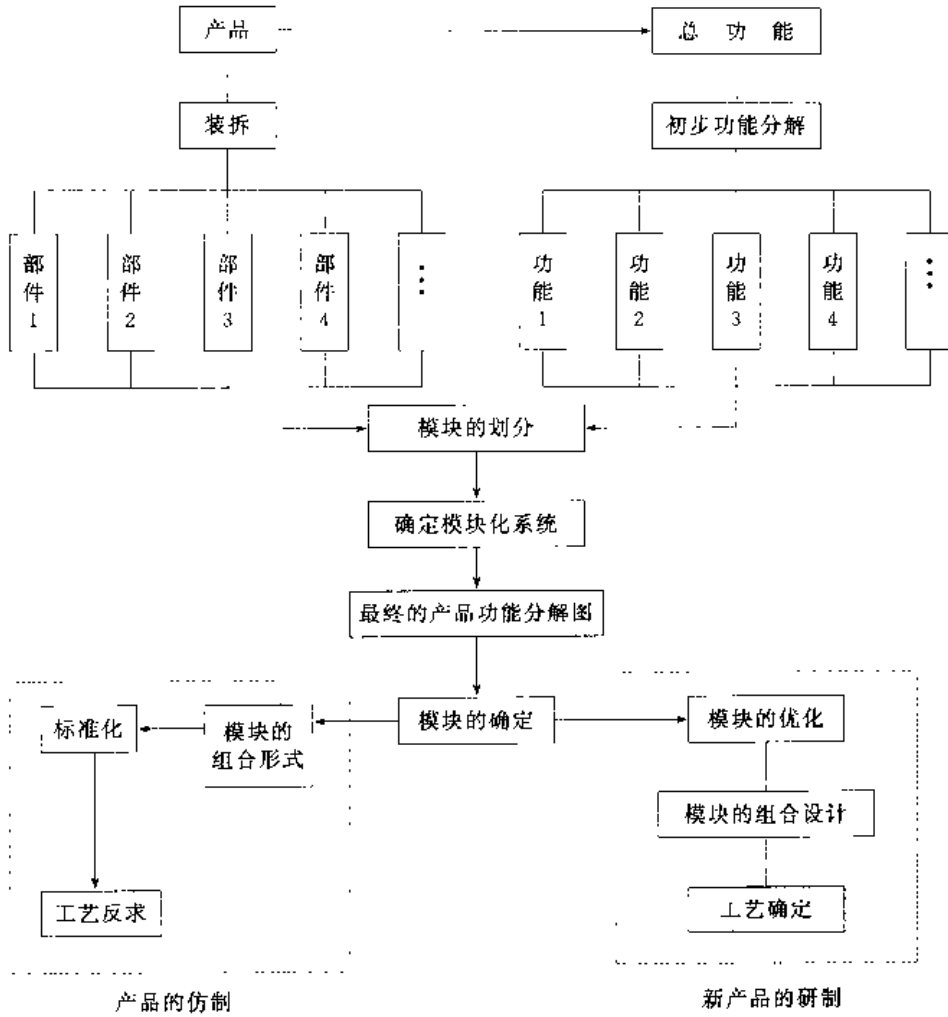


图 43.3-1 反求工程中模块划分与组合的作用

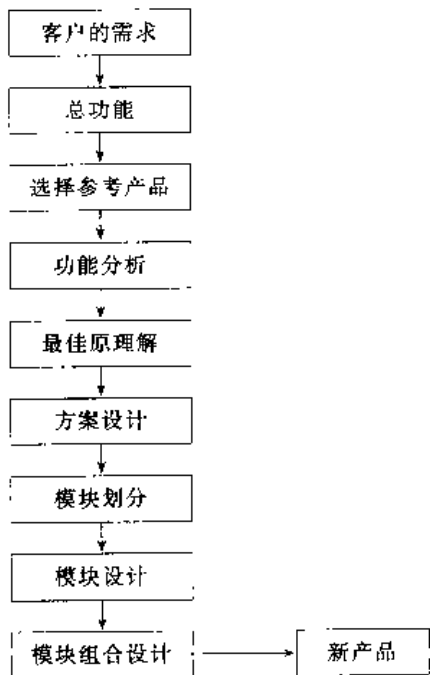


图 43.3-2 在新产品开发中模块划分与组合的作用

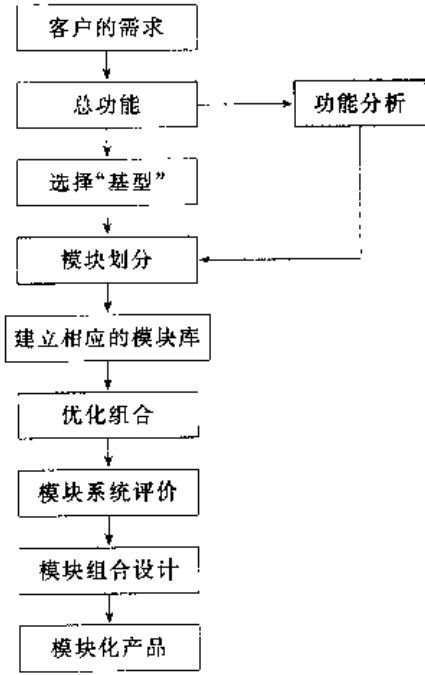


图 43.3-3 在变型产品设计中模块划分与组合的作用

件是大型零件,其生产周期,要影响产品的开发速度,故可采用同一类型建立模块。如磨床以500mm为基型,以两端的加长或压缩来建立350mm、750mm、1000mm等大件模块。

3) 模块的稳定性。在划分模块时,要注意这些模块及组成模块的基件的稳定性,即这些基件变化应是较少的。因随着某些科学技术的发展,某些结构将处于淘汰,直接影响到模块的重建,模块的易变对生产、管理等均是不利的。为此可将变化可能较大的基件单独组成模块以适应新的变化。

4) 模块的经济性。在进行系统抽象时,通过构思要建立经济合理的模块。模块的经济性主要体现在模块的通用性上,可通过大批量生产降低生产成本。即使对一些专用性的模块,批量小但由于其形状结构简单,加工方便,同样能体现出模块的经济性。

上述的模块划分原则不仅适用于反求工程,同样也适用于新产品及老产品变型产品的开发。详见图43.3-1、图43.3-2、图43.3-3。

在实际工业生产中,当以某一产品为“基型”或“参考产品”划分模块时,不可能把产品中所有的零件均划在模块中,但这些零件同样隶属于模块系统中,这种系统称之为混合系统。这种系统所构成的产品比纯模块化系统的产品要多得多,而且更实用。

### 3 模块的组合原则

人们往往希望产品在模块化后能构成数量更多的模块系统,当然有的生产单位只希望能有一定数量的不同型式的模块系统,这涉及到生产规模、产品的生命周期、同期产品的发展趋势等问题,但具体方法就与组合原则有关。例如在车床上通过附加或调整特殊功能模块使之具有钻、铣、插等其他功能,这种组合是可行的,因为它们与车削的总体精度较为一致(或低一些),但在车床上增加磨削功能,利用原有车床精度将达不到磨削精度的要求,那么这样的组合就毫无意义。因此组合并不是模块的盲目、简单叠加,而是要遵循一定的原则,以此确定哪些是能组合的,哪些是不能组合的,如何组合才能保证达到原定的要求。

模块系统(指纯模块系统和混合系统)一般在组合时将遵循三大原则:

#### (1) 目的性原则

应按具体的功能要求把它分解为基本功能和附加功能,从而组成相应的基本功能模块和附加功能模块,再通过辅助功能模块和调整功能模块的组合来体现总功能要求。凡是与总功能无关的其他特殊功能就不必组成模块。总之,模块系统的建立应以总功能作为其组合的目标。

#### (2) 灵活性原则

模块的组合应具有较大的灵活性。这种灵活性主要体现模块的功能和接触界面的互换性。

#### (3) 经济性原则

对体现功能的模块系统是采用纯模块系统还是混合系统,如何组合最有利均要从经济性角度上考虑。

## 4 模块的接口设计

模块的接口是模块的重要特性,也是模块设计的重要组成部分。由于模块的互换性是系统互换而不是单件互换,某一模块与另一模块相互的联系,实际上是某一分系统与分系统之间的关系,这种联系,借用计算机中的术语而言就是接口,即除了相关的联系尺寸、形状和接口表面的具体要求之外,尚有其他重要的参数。如录像带与录像机就存在接口问题,除了大小、厚度以外,录像带上还有起始和终止信号,还有录像机的制式问题,录像机的三种制式:日本、中国香港、西欧。否则放出来图像就非常模糊,那么制式也是接口的一个重要方面。因此,所谓模块的接口不要单纯地把它看作是连接处的相关因素,而更主要的是涉及到内部特征参数的相互匹配,模块的接口应包括两个内容:

#### (1) 相互连接的模块界面

这里所指的界面包括了接触面的形状、大小、方向、位置以及相关的连接尺寸,界面的表面特性也是很重要的部分,如各类仪器中的接插件规格,机床床身的燕尾导轨,方型平导轨等。在管系管路系统之间的连接是否采用法兰等均是界面问题。这种界面为了能使各种不同性能的模块均能任意组合,就需要使界面标准化提供给设计者,在模块设计时,根据不同的要求予以选择。尽管模块可看作是系统,但在接口界面处完全可以利用现有工艺结构的有关标准或某些标准件作为依据用到模块化设计中来。

#### (2) 模块连接时特性参数的一致性

两个模块连接是系统中的一个组成部分,每一系统中模块均可构成一种回路,有些模块是逻辑关系,但也有一些模块是物理关系或数学关系。要使这些关系能充分发挥作用,其间必存在着许多相互匹配的关联即参数的一致性,例如电气系统(如图43.3-4所示)由电气回路I与电气回路II组成,则分别为模块I与模块II,并成串联结构,模块I的输出应是模块II的输入,其间用标准的插头连接,为了保证电气回路的输出能够符合预定功能的要求,就必须要使电气回路I的输出满足电气回路II的输入要求

$$\begin{aligned} \text{其} \quad & I_1 = I_2 \\ & U_1 = U_2 \\ & R_1 = R_2 \end{aligned}$$

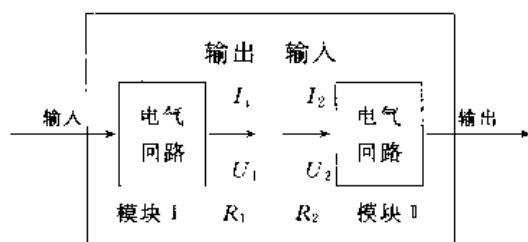


图 43.3-4 电气系统参数的一致性

又如计算机内部的存储信息,当输出与其相连时,不但是连接处的结构、形状、大小,而且更主要的是计算机的输出与工作机构的输入应该是匹配的,即工作机构的输入要求与计算机输出要求的一致性。如果各种工作机构都具有同一输入要求,那么该计算机的输出就具有较强的通用性,即输出不但可以接打印机,也可以接绘图机,也可以接其他输出系统的输出。而计算机的输入可以是键盘,也可以是 MOS,但不论何种形式,它们所给予的信号操作与计算机自身接收装置的特性也是一致的。

对机械产品来说,除了两模块连接处的结构、形状、大小之外,同样也存在模块与模块之间的参数的匹配和协调的问题。特别是对机电一体化产品来说更为重要。例如,两个不同模块上的气缸,其工作压力均应相同(除特殊要求);如工作压力不同,其中又没有任何其他调压装置,那么就无法使这两个气缸构成所说的逻辑关系。再如,在机械系统中两个零件装配后间隙等各方面均需满足要求,但由于两零件表面硬度的不一致而使其中某一零件过早的磨损,进而影响到整个产品的寿命。在发动机高温工作时,两个齿轮因材料选择不当而咬死的现象时有发生,这些均是匹配的问题。因此内部的匹配在日常生活和社会环境中是相当普遍的。产品的价格与人民生活水平也要匹配。尽管该产品是能满足需要,但购买力的影响也可能造成望“货”兴叹的情况。

为此,在模块化设计时不要将着眼点仅放在自身模块的结构、形状等方面上,还须考虑整个系统的总参数以及与相关模块的匹配。

由于模块化设计的产生使互换性又有了新的含义,即功能互换,不仅对零件适用,同样也适用于系统之间的互换。

## 5 模块化与标准化

标准化比起模块化来说,具有更悠久的历史,标准化是对重复事物和概念通过制定标准、实施标准来达到统一。在实践中,标准化的形式是多种多样的,但简化、通用化、系列化、组合作是标准化的基本形式。模块化与这些基本形式有着许多相似之处,相互交叉、相互渗透。

可以说,模块化与标准化的关系是相互依存、相互促进的。这种关系可从下面几个方面予以体现:

1) 由模块化的定义可知,模块化的产品必是以组合化为基础,尽管每一模块具有独立的功能,但这些功能仍是总功能的一部分,如果模块不进行组合,则总功能就无法实现。过去我们在标准化中应用组合化的形式时比较强调单件与单件的,这种组合一般可以认为是叠加。但模块化的出现,使组合化得到了进一步的扩展,这种组合不是简单的叠加而是通过某一联接形式进行组合。而这种组合化不是单件的组合,应是系统与系统的组合,因为模块的本身可以看作是一个子系统。组合家具只能说是应用了组合化原理,而不能说是组合家具中的某些单件是模块。

2) 模块化与系列化是使产品合理化的两条途径。所谓系列化是利用相同的求解原理,以实现在同样功能的情况下只有不同技术参数、不同结构的一组产品,它可以是横系列,纵系列,甚至跨系列进行系列化。但模块化设计的主要目的是为了开发具有多种功能的不同产品,它的出发点不是某一产品,而是一组具有多种功能的不同产品,即着眼点在于一组不同功能的不同产品。由此可见,模块化与系列化是有区别的。如果模块化产品也同时要求形成系列,那就能组成系列化的模块化产品,其通用性、适应性会更强。这里必须指出,模块并不一定都是标准化的。对于那些基本功能模块由于是反复经常出现,可以作为标准化的对象。但是对于那些特殊模块,它就不一定是标准化的对象。模块的标准化有一个完善的过程。

3) 模块化为通用化开拓前景。以前,通用化一般主要是局限于零件或部件。但模块化的产品将其模块看作是系统,那么这种通用化就成为系统通用化。如电子线路板就可以看作是模块,其具有一定的功能,这种功能可能在电子仪器中是共性的,那么该电子线路板就可以加大使用范围,这就是通用化。它对系统的维修、互换提供了很大的方便。

因此,对模块化系统的产品来说,它将通过模块标准化以达到产品标准化的目的。这就是从单元标准化向整体标准化转化的有力例证。

在模块确定以后,为了能充分利用模块的组合可能性,其关键在于保证组件的一致性以及配合面的成对性,这就要通过标准化来实现。因此:

1) 模块标准化和互换性是模块化应用的前提。

2) 模块化将为标准化开拓了新的领域,同时也对标准化提出了更高的要求,使标准化更具有针对性。

3) 模块标准化为产品的高附加值提供了有效的手段。



## 第 4 章 模块化产品的设计过程

### 1 设计过程

模块化产品的设计比一般产品的设计要复杂一些,因其产品不是按零部件分别进行的,而是针对功能进行模块的划分,其中有功能的合并及分解以及模块的匹配,使每一个模块能发挥更多的功能。

模块化产品的设计过程可用图 43.4-1 所示的流程图表示,模块化产品的设计过程不仅用于反求工程中,同时也应用于系列产品设计、新产品设计及变型产品设计,在设计时必须注意:

#### (1) 市场信息的调查、分析、处理

这是模块化产品设计的主要依据,把最终的信息处理结果汇总为需求分析汇总表。

#### (2) 做好产品规划

产品的发展规划是模块化设计的导向。

#### (3) 做好功能分解及模块的划分

以“基型”产品为参考物,结合需求进行功能分解及模块的划分。

模块的划分有一个不断完善的过程,一般均是从零部件或装拆等角度去考虑模块的划分,然后以功能的分解为依据对原有的零、部件进行归类,使之成为真正符合要求的模块。

#### (4) 建立动态的模块库

由于新技术的不断涌现,因此不断地以先进技术开发新模块去补充模块库,以丰富模块的内容,对模块库进行计算机管理,并编制相应的技术文件,为检索方便。模块库中的模块不但有类别上(功能的)区别,也有层次上的区别。例如,按通用功能、专用功能分大类,通用功能又可分为八大类,如增大、缩小、闭合、分离等,同一功能下又有不同的结构。在技术管理中一定要达到这样的要求:点击一个结构就能给出一个相应的技术与之相对应。

#### (5) 建好模块的接口形式库

模块保证其互换性,接口的形式、大小等均是非常重要的,因此模块的接口形式库的标准化将会大大缩短模块化设计的周期。

### 2 需求分析

在模块化产品的设计过程中(如图 43.4-1 所示),需求分析是设计的基础,需求分析应包括如下的内容:

1) 不管是老产品,还是引进产品,还是希望开发

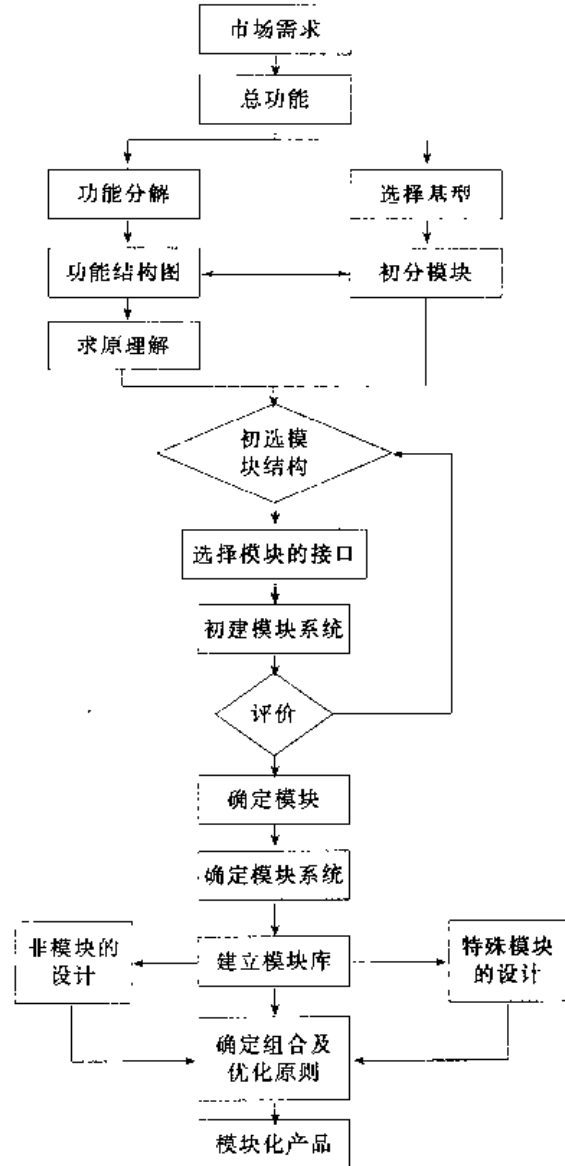


图 43.4-1 模块化产品的设计过程

的新产品,其功能是否满足了市场的需要,哪一些功能是必须的就属于基本功能,而哪一些功能使用较少的,则以特殊功能处理。

2) 对功能进行分类以观察其变化的状况,这将涉及到开发的模块的数量,变化大的,模块数量要多。

3) 客户对功能载体所采用的结构及其特性的态度。

4) 同类产品对市场占有率的分布状况的原因分析。

5) 客户对产品性价比的看法。

6) 产品的发展趋势及展望。

7) 预计产品的需求量及产品生命周期。

### 3 可重构设计的概念和模型

可重构设计的基本思想是根据顾客对产品的需求,在充分重用已有设计经验的基础上通过选择已有的预定义的构件,以及展开新构件设计活动来快速重构满足重构约束的目标产品。可重构设计系统由六个模块(设计资源模块、顾客需求定义模块、产品功能结构映射模块、部件/特征件抽取模块、过程件抽取模块、

设计控制模块)构成,各个模块在一定的规则下,通过既定的方式与其他模块发生作用。

可重构设计模型(如图 43.4-2 所示)中各个流程的顺利进行不仅依赖于上述各个完整的模块,还依赖于各个模块之间的有效的相互作用。从用户需求分析、构件选择到设计过程规划都需要参照集成化配置本体模型。

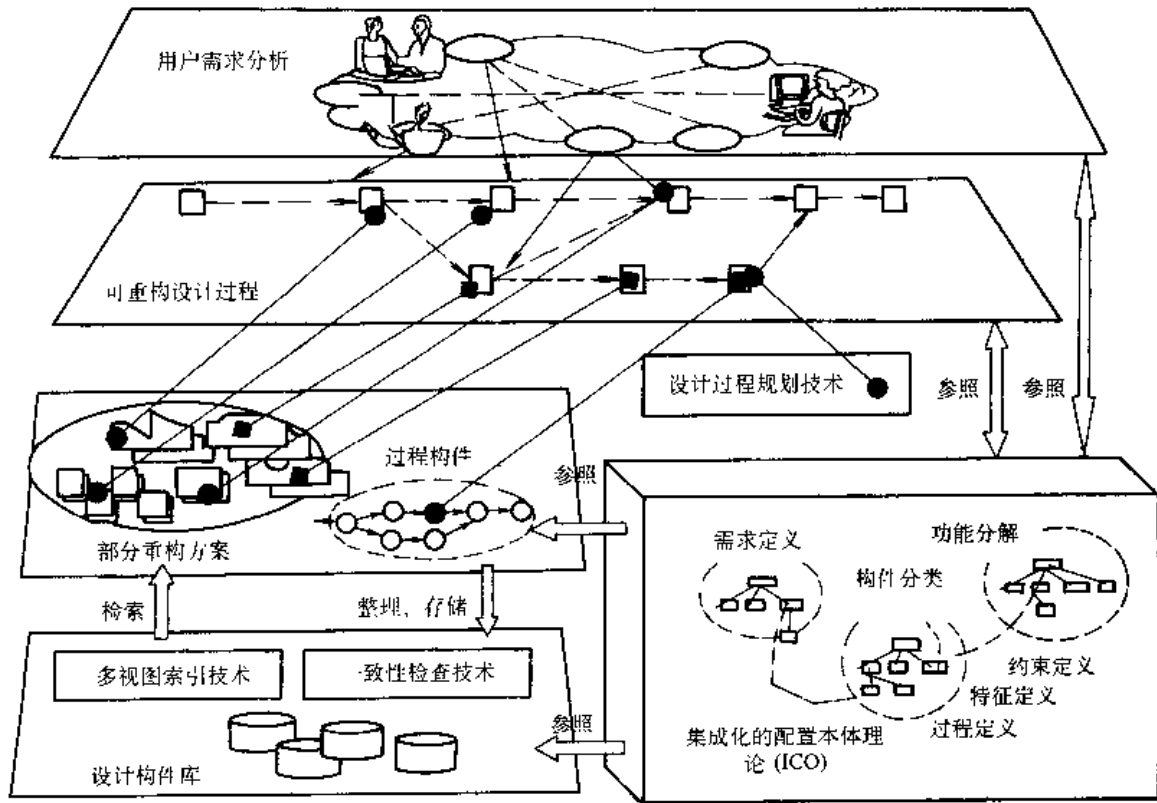


图 43.4-2 可重构设计过程流程示意

### 4 设计中的单元重用

选择和重用已有零件设计方案是复杂产品创新的基础。重用以往设计方案,快速合理地响应顾客的个性化需求是模块化设计的基本目标之一。

#### 4.1 重构单元的概念和分类

一般配置设计中,构件(component)包括预定义(predefined)的所有的零件(piece part)、部件(assembly part)的集合。

根据 Mittal 的定义,每一个构件被描述为一组属性相关且与其他零件联结的接口(ports),以及每一接口处需联结的构件的约束(constraints)和其他结构约束。

早期的配置多用于计算机软硬件系统。在公开发表的大量配置文献中,大都采用计算机硬件作为分析实例。由此,也可以看到研究者对这类系统的偏爱。在

这类系统中,较强的可配置性是一个显著特征,形式简单、整齐划一的接口类型也接近理想化的配置问题描述。例如,在软件配置中,部件指程序模块;接口指模块间的传递参数;约束定义了各参数赋值的数目和类型,或者是关于程序模块间协调的一些约束描述;关于理想配置的描述是指用户对其编写软件的期望说明。

Mittal 对其定义的配置还做了如下三点重要的限制,即:

- 1) 在配置设计任务中,不得出现新部件(new component);
- 2) 每一个预先定义好的部件都必须以“预定义的”方式联结在一起,不得对联结方式做修改而得到任意的联结。
- 3) 问题的解不仅要列出所选部件的清单,而且应当指出其联结方式。

通过上面的三条限制可以看出,此类研究中配置

问题关于构件的定义是非常理想化的。

在可重构设计中,构件是指已有的设计知识的单元,是知识重用的单元。具体地说,它是零(部)构件、特征构件和过程构件的统称。这里用大写英文单词 PART 来表示构件。如图 43.4-3 所示。

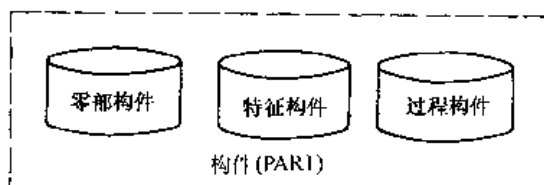


图 43.4-3 可重构构件内容

其中,零(部)构件、特征构件和过程构件又可以分别简称为:部件、特征件和过程件。为了描述方便,除了用大写英文单词 PART 来表示构件外;部件则用 component 来表示,特征件与过程件分别用 feature part 与 process part 来表示。

#### (1) 零部构件

零部构件简称部件(component)。部件包含零件(picce part),或称原子部件(primitive component),和装配件(assembly part),或称复合部件(composite component)。部件是基本的重构构件,这里我们将重点阐述另两类较新的构件内容:特征构件和过程构件。

#### (2) 特征构件

特征构件简称特征件(feature part)。特征的概念最初源于工艺设计,指形状特征,如:退刀槽特征、倒角特征等。随着特征研究的深入,特征的定义也得到了极大的拓展,如原型特征、或产品全生命周期信息特征等。本文的特征指这种经过发展的广义的设计特征,但处理方式上将与其他研究有所变化。人们通常将特征看作是零部件的属性来研究特征。这一点反映了特征与零部件之间存在的依附关系,即在一个产品中并没有脱离零件而独立存在特征,这是毫无疑问的。在本文的研究中,我们将特征看作一个抽象的构件,并不是说在最终的设计方案中它是独立于具体零部件存在的,而是提出在知识存储、管理、问题求解、方案配置的过程中,它可以相对独立地处理,也就是说,在可重构的设计中,将特征看作是一个相对独立的设计元素,看作是一个抽象的构件,通过其明确的“接口需求”定义,可以方便地与零部件“装配”在一起。

#### (3) 过程构件

过程构件简称过程件(process part)。由于现实产品的复杂性和我们知识占有的有限性,在给定的求解目标下,当完成了一次对仅含有部件和特征件的知识库的搜索后,经常只能得到一个最大化的部分配置解(maximum partial configuration),我们把这个过程称

为初步重构(primary configuration)。在这种情况下,为了构造出目标产品,还需要展开新的设计开发子项目。初步重构阶段的一些遗留信息,将为辅助新设计项目的展开提供依据。在可重构的设计过程中,我们试图展开一个后续的构件选择过程,即过程件选择。从这一角度出发,过程构件将被视为构件的第三部分内容。一个系统的过程构件重用的程度取决于企业拥有过程资源的种类和数量,以及这些过程构件资源的知识所表示的内容。

前面提到了“设计知识重用(reuse)”的概念,那么设计知识重用会带来那些好处呢?

#### (1) 提高设计响应速度

零部件的重用是众多设计系统已经实施的做法。对某个设计领域来讲,尤其针对某个产品系列来讲,如果我们能将企业或行业出现的零部件的设计知识和一些设计特征(包括关于该产品的细节造型,某些零部件的改型,以及针对特定顾客个性需求形成的特殊解决方案)组织、存储、管理起来,丰富设计知识库。在新的顾客需求来到时,通过检索这类“资料”,并将其重用,从而加快设计进程。过程构件则是针对展开的新设计,从过程管理、规划的角度去重用资源、加快设计进程。

#### (2) 降低设计成本、提高质量

从通常的产品设计活动角度来看,功能最终是由一系列的零部件组合而成的结构来实现的。当新的顾客需求(经常表现为产品辅助需求)出现后,经过需求一功能映射,系统可能做出以下三种处理:①在原配置方案的基础上,选择某个兼有该辅助功能的零部件替换相应部件;②在原配置方案上增加一个能够实现该辅助功能的零部件;③重新设计一个与原方案可装配的实现新功能的零件。

从设计成本的角度来说,三个策略中①较理想,其余两方案(尤其③)成本较大。然而,实际情况中,现有的零部件库中往往不一定能够找到符合①中所述条件的零部件。而开发新的零部件代价又过大。

零部件的重用本身避免了零部件数目的无限扩大,设计人员不再依靠直觉来随意设计“新”的零件。这是最直接的好处,除此之外,还会因此为设计知识库、实际库存、生产装备等若干环节带来间接的益处。

此外,特征件的引入为零部件提供了设计差异的空间,特征件与部件的多种组合为简化设计方案中的总零件数目带来好处。

## 4.2 单元重构过程

### 4.2.1 RPD 概念

在产品设计中,一方面,很多供应商都设定了这样

的目标,即比较高的灵活性尽可能满足每一位顾客提出的特定要求。为了达到这个目标,产品设计部门尽其所能,以期设计出能够满足各种各样特殊要求的完全方案。于是,开发人员依靠直觉不断重复“发明”新的零件,而全然不顾新的零件与传统零件的大同小异。与此同时,与订单相关的零件和物品编号的数量也无限制地增加,从而增加了固定成本,也影响了产品的交货期和质量。作为对策,产品标准化正是解决上述问题的有效途径之一。标准化并不是创新的反义词,而是以“减少产品全生命周期各阶段的数据冗余,为创新理清思路,减少由信息不一致而导致的成本增加和质量不稳定”为目标的。另一方面,文献表明,在创新设计中,产品中大多数的零件或多或少有其历史原型,而相对来讲,全新的零件仅占很少一部分。甚至,有文献认为“所谓创新性,是指对过去经验和知识的分解组合,使之实现新的功用”,“企业的创新能力是其知识量与扩散思维的综合”。因此,从两方面来看,重用历史设计经验具有重要意义。

本章提出的面向创新的可重构产品设计(Reconfigurable Product Design,RPD)技术正是为了从设计角度实现产品的标准化,以创新为目标利用设计经验而采取的一种设计方法。可重构产品设计(RPD)的含义是指设计人员通过继承并且重构历史的设计方案面进行一项产品创新设计的方法。作为一种产品设计方法,可重构的产品设计方法是以“充分、有效利用设计经验”为基本思想的。

该方法具有以下特征:首先,它强调在检索历史零件设计方案的过程中注重零件的“可重构性”,即强调对待选择的零件之间的“匹配性”问题。它寻找的是一种更加合理的全设计过程的全局最优方案,而不是仅仅分别地满足功能、性能约束的零件个体约束的局部的最优解。因此,RPD方法是以全局为出发点的。其次,在基于零件重构约束的零件单元的搜索之中,RPD方法采用的是一种优化的“启发式”搜索方法,从而在降低搜索空间的同时,保证了后续的待选择零件的“可选择域”。因此,RPD方法是一种与人工智能相结合的设计方法。

#### 4.2.2 RPD 流程

从本质上讲,RPD方法是对传统的基于实例的设计思想(Case Based Design,CBD)(如图 43.4-4 所示)的一种新的推广。与传统的应用相比,在检索零件的过程中(如图 43.4-5 所示),首先强调了整体约束及其动态传播;其次是检索策略的优化。在 4.4.3 中将详细讨论这两点问题。

在设计实际应用中,现有的 CBD 的方法通常没有

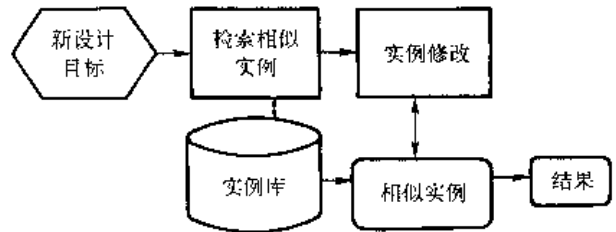


图 43.4-4 基于实例的设计流程

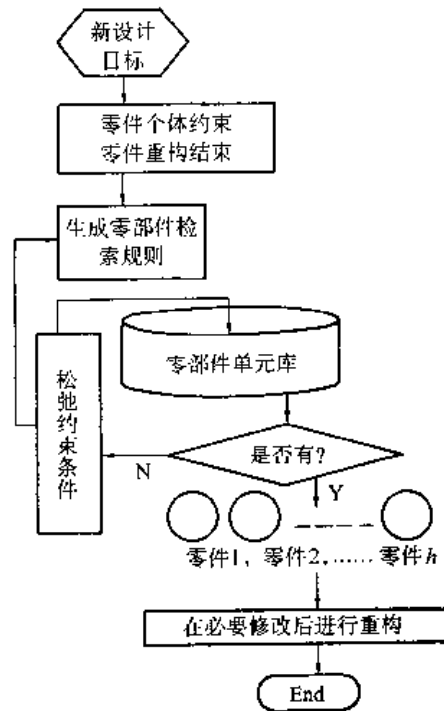


图 43.4-5 RPD 方法构件选择流程图

考虑零件单元之间的重构约束条件。从 RPD 方法流程图可以发现,它较好地解决了这一问题。首先,该方法以“单元”的角度去考虑待设计实例,采用了从零件个体角度(功能、性能)约束与从产品整体角度约束(零件的可重构性)相结合的检索机制,其中后者是一种随着零件单元的不断选出而“动态变化”的约束,如待选取零件和已选取零件之间的可装配性约束,所选取的所有零件重量的总和约束等。

#### 4.2.3 RPD 关键技术

按照构造一个应用系统的先后次序,可以将关键技术粗分为以下三大类。

##### (1) 合理的零件单元的存储技术

包括如何优化历史的零件设计单元的存储方法和数据库内部结构。优化的准则包括以下三点:第一,能够在存储阶段为以后的零件单元的检索和选择预先考虑。所采用的存储机制应当是从零件功能约束、性能约束,以及重构性约束等角度对零件单元库进行检索提

供可能与方便。第二,采取的存储机制应当具有开放式的数据结构,允许新的零件单元不断进入单元库。第三,在一个新的零件进入数据库之前,存储机制能够提供必要的检查,归并程序,防止单元数据的“冗余性”和“不一致性”。

#### (2) 有效的零件单元的选择技术

其包括如何有效地表达零件单元的功能、性能约束条件,尤其是零件单元之间的“重构约束”条件;如何构建优化的搜索策略;如何松弛或变更解的搜索条件。

#### (3) 有效的单元修改方法

经过选择满足个体功能、性能,以及重构约束的相似单元,可以重构出“初步设计目标”。然后再对该“初步设计目标”进行必要的修改,即得到最终设计目标。

下面将主要针对其中的零件单元的选择技术做较深入地探讨。

### 4.3 优化的零件单元选择技术

零件的选择过程是在满足一定约束条件下对历史零部件单元库进行搜索的过程。其中的约束条件应该由以下两部分构成。第一,零件个体属性约束;第二,零件重构约束。下面给出这两种约束的定义。

#### (1) 零件个体属性约束

在零件选择过程中,零件个体属性约束是一组相互独立的约束的集合,该集合中的每一个元素反映了从个体的某种功能、性能的属性角度出发对所选择零件的约束。

#### (2) 零件重构约束

在零件选择过程中,设计方案中各个零件之间应当满足的约束称为零件重构约束。假定有  $n$  个零件  $p_1, p_2, \dots, p_n$  待选择,零件重构约束一般可记为  $C_i(p_1, p_2, \dots, p_n), (i=1, 2, \dots, m)$ 。多个零件重构约束构成零件重构约束的集合。零件重构约束以下简称重构约束。

零件单元的选择是一个复杂的约束满足问题,从理论上讲,它属于 NP 问题。假定有 10 个零件待选取,每一种零件平均有 6 种类型,则穷尽搜索的复杂度是  $10^6$ 。当待选择的零部件数目增多,每一个零件的可选域增加时,问题的复杂度将更大,因此,寻找一种好的搜索策略显得十分重要。

#### 4.3.1 零件选择策略

随着零件逐渐被选出,零件重构约束的内容将具体化。实际上,越靠后被选出的零件所受到的重构约束条件将越加严格。或者说,先取出的零件将对在其后取出的零件的可选域带来损失(range lost)。给定一个

零件选择的顺序,或在选取某类零件时,选择不同的零件,就将暗含了一定的选择范围损失总和(total range lost)。同样,较大的选择范围损失无疑将错失好的零件选择,不利于产品重构目标的实现。

本文提出了一种零件选择策略,该策略以有效的启发式方法降低零件选取过程的复杂度;并在此基础上通过构建适合的零件类选择顺序和零件簇成员选择准则,来尽可能地缩小选择范围损失总和,从而最终得到满足约束条件的零件簇。

#### 4.3.2 最强约束策略和最小损失策略

在零件选择过程中,可采用两个主要的策略,一是最强约束策略,与此策略对应的是一个“关键零件函数  $F$ ”,该函数总是计算出当前情况下,可选零件的数目最少的零件类,该零件类被视为最强约束零件类。在决策过程中,将首先从最强约束零件类中确定零件。这一策略的一个最简单情况是,某一零件类中只有一种零件可供选择,此时首先考虑这类零件是十分自然的。最强约束零件类的灵活性最小,不能被延误,因此具有最高优先级。第二个策略是最小损失策略,即在具体选择零件时,应保证其他待选取零件的可选空间保持最大,即对其他待选零件的选取产生最小的影响。这种影响的程度通过“可选域损失函数  $L$ ”来度量,用来计算当选定了当前零件时,其他零件类的有效解的数目的递减。

在以下的算法中,我们将首先选出符合零件个体约束的各类零件。然后根据关键零件函数  $F$  的计算,来确定关键零件类。在某类零件中选择零件时,根据可选域损失函数  $L$  的计算,来确定最有希望的零件个体,因此函数  $F$  以及函数  $L$  的计算是关键。

$F$  是定义在每一个待选取零件类上的,它表示在当前的重构约束下,不明显违背约束的该类零件的个数。所谓不明显违背约束是指,选择该类零件中的某零件不直接违背重构性约束,尽管可能最终导致无解。而  $L$  则是定义在当前待选零件类中的每一零件上的,它表示选择该零件后,对其他零件类选择域大小的度量,假定当前零件类为  $p_k$ ,其他待选零件类分别为  $p_{k+1}, p_{k+2}, \dots, p_n$ 。假定在第  $k$  类零件选择前,  $F(p_{k-1}) = f_{k-1}, F(p_{k+2}) = f_{k+2}, \dots, F(p_n) = f_n$ 。并假定此时,在第  $k$  类零件中,有  $r$  个被选零件。如果在选择第  $k$  类零件中的第  $j$  个零件  $p_{k,j}$  后,  $F(p_{k+1}^j) = f_{k+1,j}, F(p_{k+2}^j) = f_{k+2,j}, \dots, F(p_n^j) = f_{n,j}$ 。则可选域空间的大小  $L(p_{k,j})$  可以用

$$f_{k+1,j} \times f_{k+2,j} \times \dots \times f_{n,j} \\ = \prod f_{i,j} (i=k+1, k+2, \dots, n).$$

$$L(p_k) = \max(L(p_{k,j})), (j=1, 2, \dots, r)$$

$L(p_k)$ 就是当前零件类  $p_k$  的可选域损失函数  $L$  的值,此时可选域的损失将最小。

随着零件的不断被选出,现有零件重构约束也将发生变化,产生新的零件重构约束,这样的过程称作零件重构约束的约束传播。约束传播的计算,针对不同的约束属性,复杂程度将有所不同。当约束属性是“重量、成本”等线性约束时,约束传播的计算比较简单。比如,某个产品由 10 个待选择零件构成,总成本约束要求不超过 600 元,当一个成本为 35 元的零件选定后,新的成本约束为:其余 9 个零件的总成本不超过  $600 - 35 = 565$  元。而当约束属性是装配约束等较复杂的约束时,约束传播的计算过程往往比较复杂。这种复杂的计算过程经常需要一些相关知识的指导。

### 4.3.3 零件选择算法

图 43.4-6 表示了零件选择算法的原理图,具体步骤如下:

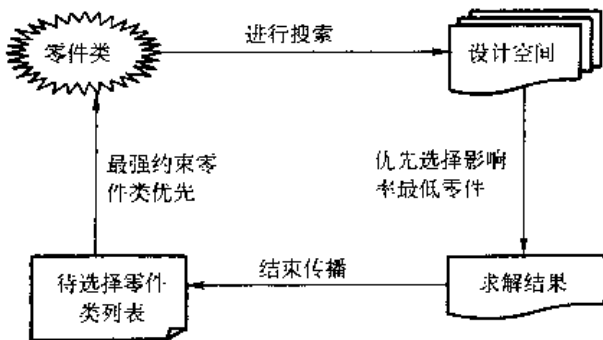


图 43.4-6 零件选择算法

- step1: 初始状态下,“待选取零件类列表”包含所有零件类。在数据库中分别筛选出符合“零件个体约束”的各个待选取零件类的集合,求出各个集合所含元素个数,根据函数  $F$  计算出关键零件类;
- step2: 以当前的“关键零件类”作为即将进行搜索的零件类  $p_k$ ;
- step3: 计算即将进行搜索的零件类的函数  $L$  的值:  $L(p_k)$ ,选取与之对应的零件  $j$  作为本零件类的选择结果,将其归入已选择零件簇,并将零件类  $p_k$  从“待选取零件类列表”中删除,得出当前“待选取零件类列表”;
- step4: 对“零件重构约束”进行“约束传播”计算,得出当前“零件重构约束”;
- step5: 根据当前“零件重构约束”,在“待选取零件类列表”上计算函数  $F$ ,求得关键零件

类;

step6: 转入 step2;

⋮

step7: 直至“待选取零件类列表”为空时,终止。

算法补充说明:①若求解过程中出现当前可选域为空的情况而无法继续搜索,则回溯到前一步重新进行求解。②符合条件的零件簇可能并不唯一。③若按此算法并结合回溯策略,经过一定计算仍无法求得满足约束解,则意味着需要适当松弛约束或扩充零件库的内容。

### 4.4 应用举例

以设计一个新的剃须刀为例,经过对该产品的分析,发现有 6 个零(部)件可以重用历史设计零件单元,仅有 1 个零件需要重新开发。现在通过不考虑重构约束、考虑重构约束采用未优化的选择方法、考虑重构约束采用优化后的启发式选择方法三种情况来对比说明 RPD 方法的优越性。为了方便地说明问题,本例中,以重量、成本作为两个子约束。

假设剃须刀总重量要求不大于 115g。材料均采用防水,绝缘材料。新开发的“刀刃”总重量不超过 5g,所以待选择的 5 个零件(如图 43.4-7 所示,从左向右分别记做  $a_1, a_2, \dots, a_5$ )总重量不得超过 110g,并要求待选零件总成本不大于 76 元。

下述内容为该设计的过程:

(1) 确定待选择零件类

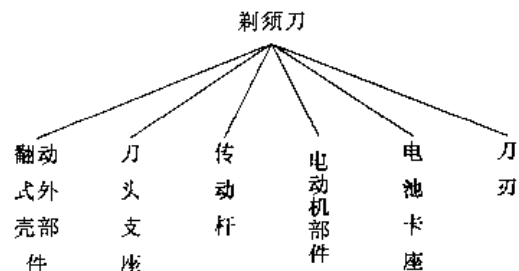


图 43.4-7 零部件分解图

(2) 选择过程

表 43.4-1 零件库中满足个体约束的零件集合

零件类	零件类的零件数	零部件单元	重量/g	成本/元
$a_1$ 外壳	8	1	35	32
		2	41	22
		3	56	12
		4	65	10
		5	79	14
		6	88	18
		7	95	10
		8	105	9

(续)

零件类	零件类的零件数	零部件单元	重量/g	成本/元
a <sub>2</sub> 刀头支座	4	1	20	5
		2	16	6
		3	8.8	7.5
		4	4.4	22
a <sub>3</sub> 传动杆	6	1	16	2
		2	12	2.5
		3	5.2	8
		4	5.5	4
		5	6	3
		6	10	3
a <sub>4</sub> 电动机部件	3	1	40	10
		2	43	37
		3	50	50
a <sub>5</sub> 电池卡座	6	1	14	0.5
		2	8	0.6
		3	6	0.8
		4	3.5	2
		5	4.5	1.2
		6	9	0.65

step1: 待选择零件列表(如表 43.4-1 所示): a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, a<sub>4</sub>, a<sub>5</sub>。

约束: C<sub>1</sub>: g(a<sub>1</sub>) + g(a<sub>2</sub>) + ... + g(a<sub>5</sub>) ≤ 110,

C<sub>2</sub>: c(a<sub>1</sub>) + c(a<sub>2</sub>) + ... + c(a<sub>5</sub>) ≤ 76。

函数 F: 由 C<sub>1</sub>, C<sub>2</sub> 得出 C<sub>1</sub>: (∀ i) g(a<sub>i</sub>) ≤ 110; C<sub>1</sub>: (∀ i) c(a<sub>i</sub>) ≤ 76; 由此计算得到 F(a<sub>1</sub>) = 8, F(a<sub>2</sub>) = 4, F(a<sub>3</sub>) = 6, F(a<sub>4</sub>) = 3, F(a<sub>5</sub>) = 6。

step2: 由于 F(a<sub>4</sub>) = 3 最小, 因此将 a<sub>4</sub> 作为关键零件类。

step3: 以下考察 a<sub>4</sub> 的三个选择 a<sub>4</sub>(1), a<sub>4</sub>(2), a<sub>4</sub>(3), 计算函数 L(p<sub>4</sub>)。

若选择 a<sub>4</sub>(1): g = 40, c = 40, 此时, 约束 C<sub>1</sub>: g(a<sub>1</sub>) + g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 70; C<sub>2</sub>: c(a<sub>1</sub>) + c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 36; 约束松弛为 g(a<sub>i</sub>) ≤ 70, c(a<sub>i</sub>) ≤ 36 (i = 1, 2, 3, 5) 据此计算 F(a<sub>1</sub>) = 4, F(a<sub>2</sub>) = 4, F(a<sub>3</sub>) = 6, F(a<sub>5</sub>) = 6, L(p<sub>4,1</sub>) = 4 × 4 × 6 × 6 = 576;

同理, 若选择 a<sub>4</sub>(2): g = 43, c = 37, 此时, 约束 C<sub>1</sub>: g(a<sub>1</sub>) + g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 67; C<sub>2</sub>: c(a<sub>1</sub>) + c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 39; 约束松弛为 g(a<sub>i</sub>) ≤ 67, c(a<sub>i</sub>) ≤ 39 (i = 1, 2, 3, 5) 据此计算 F(a<sub>1</sub>) = 4, F(a<sub>2</sub>) = 4, F(a<sub>3</sub>) = 6, F(a<sub>5</sub>) = 6, L(p<sub>4,1</sub>) = 4 × 4 × 6 × 6 = 576;

若选择 a<sub>4</sub>(3): g = 50, c = 50, 此时, 约束 C<sub>1</sub>: g(a<sub>1</sub>) + g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 60; C<sub>2</sub>: c(a<sub>1</sub>) +

c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 26; 约束松弛为 g(a<sub>i</sub>) ≤ 60, c(a<sub>i</sub>) ≤ 26 (i = 1, 2, 3, 5) 据此计算 F(a<sub>1</sub>) = 2, F(a<sub>2</sub>) = 4, F(a<sub>3</sub>) = 6, F(a<sub>5</sub>) = 6, L(p<sub>4,1</sub>) = 2 × 4 × 6 × 6 = 288;

因此 L(p<sub>4</sub>) = max(L(p<sub>4,j</sub>)) = 576, 取最大值时 j = 1 或 j = 2, 注意到此时 a<sub>4,1</sub>, a<sub>4,2</sub> 都可作为可选零件, 可以从中任选其一。结合算法补充说明(1), 当所选的那个零件导致后续求解过程中出现当前可选域为空的情况而无法继续搜索时, 则回溯到该点, 选择另一个零件重新进行求解。在这里, 试选择 a<sub>4,2</sub> (43, 37)。

待选取零件列表变为: a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, a<sub>5</sub>。

step4: 约束传播为(详见 Step3 中相关步骤):

C<sub>1</sub>: g(a<sub>1</sub>) + g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 67,

C<sub>2</sub>: c(a<sub>1</sub>) + c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 39;

step5: 当前 F 函数值为: F(a<sub>1</sub>) = 4, F(a<sub>2</sub>) = 4, F(a<sub>3</sub>) = 6, F(a<sub>5</sub>) = 6。故下一步选择 a<sub>1</sub> 为当前的关键零件类。

step6: 转入 step2, 分别考察 a<sub>1</sub> 的 4 个零件 a<sub>1</sub>(1), a<sub>1</sub>(2), a<sub>1</sub>(3), a<sub>1</sub>(4), 计算函数 L(p<sub>1</sub>)。

第一个零件 g(a<sub>1(1)</sub>) = 35, c(a<sub>1(1)</sub>) = 32, 约束变为 C<sub>1</sub>: g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 32; C<sub>2</sub>: c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 7。约束松弛为 g(a<sub>i</sub>) ≤ 32, c(a<sub>i</sub>) < 7 (i = 2, 3, 5) 据此计算 F(a<sub>2</sub>) = 2, F(a<sub>3</sub>) = 5, F(a<sub>5</sub>) = 6。L(p<sub>1,1</sub>) = 2 × 5 × 6 = 60。第二个零件 g(a<sub>1(2)</sub>) = 41, c(a<sub>1(2)</sub>) = 22, 约束变为 C<sub>1</sub>: g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 26; C<sub>2</sub>: c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 17。约束松弛为 g(a<sub>i</sub>) ≤ 26, c(a<sub>i</sub>) ≤ 17 (i = 2, 3, 5) 据此计算 F(a<sub>2</sub>) = 3, F(a<sub>3</sub>) = 6, F(a<sub>5</sub>) = 6。L(p<sub>1,2</sub>) = 3 × 6 × 6 = 108。第三个零件 g(a<sub>1(3)</sub>) = 56, c(a<sub>1(3)</sub>) = 12, 约束变为 C<sub>1</sub>: g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 11; C<sub>2</sub>: c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 27。约束松弛为 g(a<sub>i</sub>) ≤ 11, c(a<sub>i</sub>) ≤ 27 (i = 2, 3, 5) 据此计算 F(a<sub>2</sub>) = 2, F(a<sub>3</sub>) = 4, F(a<sub>5</sub>) = 5。L(p<sub>1,3</sub>) = 2 × 4 × 5 = 40。第四个零件 g(a<sub>1(4)</sub>) = 65, c(a<sub>1(4)</sub>) = 10, 约束变为 C<sub>1</sub>: g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 2; C<sub>2</sub>: c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 29。约束松弛为 g(a<sub>i</sub>) ≤ 2, c(a<sub>i</sub>) ≤ 29 (i = 2, 3, 5) 据此计算 F(a<sub>2</sub>) = 0, F(a<sub>3</sub>) = 0, F(a<sub>5</sub>) = 0。L(p<sub>1,4</sub>) = 0 × 0 × 0 = 0。

L(p<sub>1</sub>) = max(L(p<sub>1,j</sub>)) = 108, 取最大值时 j = 2, 选择 a<sub>1,2</sub> (41, 22)。

约束传播为

C<sub>1</sub>: g(a<sub>2</sub>) + g(a<sub>3</sub>) + g(a<sub>5</sub>) ≤ 26,

C<sub>2</sub>: c(a<sub>2</sub>) + c(a<sub>3</sub>) + c(a<sub>5</sub>) ≤ 17。

约束松弛为 g(a<sub>i</sub>) ≤ 22, c(a<sub>i</sub>) ≤ 17 (i = 2, 3,

5) 据此计算  $F(a_2)=3, F(a_3)=6, F(a_4)=6$ 。选择  $a_2$  为关键零件类。

待选取零件列表变为:  $a_1, a_3, a_4$ 。

∴  
∴

以下同理: 待选取零件列表变为:  $a_3, a_4$ 。

∴  
∴

待选取零件列表变为:  $a_5$ 。

∴  
∴

待选取零件列表变为空, 终止。选中的一组零件单元:  $a_{1,2}(43, 37), a_{1,2}(41, 22), a_{7,3}(8, 8, 7, 5), a_{1,4}(5, 5, 4), a_{5,3}(6, 0, 8), G=104.3 < 110(g), C=71.3 < 75(\text{元})$

### (3) 结果比较

作为对比, 表 43.4-2 中的选择结果是不考虑零件间的重构约束下所得出的:

表 43.4-2 不考虑重构约束的一组选择结果

各部件名称	待选零件类大小	选择结果(g, 元)
$a_1$	8	$a_{1,2}(41, 22)$
$a_2$	4	$a_{2,2}(16, 6)$
$a_3$	6	$a_{3,6}(5, 2, 8)$
$a_4$	3	$a_{4,1}(40, 40)$
$a_5$	6	$a_{5,5}(9, 0, 65)$

表 43.4-3 中的抽取结果是考虑零件间的重构约束, 但采用未经优化的选取过程所得的结果为:

表 43.4-3 考虑重构约束未优化选择过程所得结果

各部件名称	动态重构约束	可抽取域	选择结果(g, 元)
$a_1$	$G(a_1, a_2, a_3, a_4, a_5) \leq 110$ $C(a_1, a_2, a_3, a_4, a_5) \leq 75$	8	$a_{1,1}(35, 32)$
$a_2$	$G(a_2, a_3, a_4, a_5) \leq 75$ $C(a_2, a_3, a_4, a_5) \leq 43$	4	$a_{2,1}(20, 5)$
$a_3$	$G(a_3, a_4, a_5) \leq 55$ $C(a_3, a_4, a_5) \leq 38$	6	$a_{3,1}(16, 2)$
$a_4$	$G(a_4, a_5) \leq 39$ $C(a_4, a_5) \leq 36$	0	找不到重量不大于 39g 并且成本不大于 36 元的电动机部件
$a_5$	—	0	—

通过表 43.4-2、3 容易看出, 表 43.4-2 中仅从局部出发, 寻求满足个体功能、性能要求, 如造型、结构、成本的最优解, 没有考虑重构约束。结果中的 6 个零件的总重量等于 111.2g, 总成本等于 76.65 元, 大大超出了初步设计目标  $G \leq 110g, C \leq 76$  元的要求。表

43.4-3 中由于考虑了重构约束, 但没有优化选取过程, 因此出现靠后选择零件可选择域过小的情况, 而无法完成设计。而考虑了重构约束, 并采取启发式的搜索过程, 选取结果的可重构性大大地提高了。

## 5 量化法在设计过程规划中的应用

由于顾客需求的差异性、产品结构的复杂性以及设计系统所占有知识的有限性, 当一个可重构设计系统完成了构件选择环节后, 并不一定能得到最终符合顾客需求的该产品的完全重构方案。在这种情况下, 一些新的设计模块需要展开。这些新展开的设计子项目和已有构件的配置过程共同构成了一个完整的可重构设计项目。如何在有限资源下成功地展开整个设计过程, 并且达到项目要求, 成为项目管理的瓶颈所在。基于量化原理的设计评审规划技术, 将能通过合理定位评审点的方法来使设计中的有限资源得到较合理的分配, 从而使设计过程的整体性能得到提高和改善。

设计进度的超期和成本超支, 长期以来困扰着设计规划人员和项目管理人员。有报告表明, 一组被调查的设计项目的进度超期达 41%~258%, 成本超支则达 97%~151%, 近二十年来, 随着设计技术(如 CAX 等)的发展和提高, 这个问题却没有得到缓解。快速响应顾客需求, 利用重构技术和知识重用手段快速交付产品是可重构设计的基本目标之一, 如何解决设计过程(包括新构件的设计和重用构件的选择)规划问题成为可重构设计继“构件选择”之后的又一关键技术。对产品设计领域的研究表明, 成本的超支会导致整个产品开发项目的失败或终止, 而进度拖延则会由于错失市场的新需求而失去市场占有率并增加产品退化的可能性。因此, 有研究人员曾断言: “delay is deadly”。可见, 如何从设计过程来保证质量, 降低成本, 减少产品上市周期, 是设计人员所关心的内容。设计过程中各活动之间会有各种潜在的联系和相互作用。有时, 某个设计任务没有按时完成, 会对位于其后的一系列任务, 甚至包括与其分享相同设计资源的其他项目, 造成多米诺(Domino)骨牌效应, 继而导致整个设计项目的拖延或失败。因此, 通过对设计流程进行有效地规划从而确保设计质量, 及早地发现问题的源头, 以改正设计中存在的问题, 显得十分重要, 尤其是对一些产品提交时间要求严格的设计项目。

关于设计项目的过程规划和管理, 目前在国际上的人多数算法所依赖的模型中, 对于设计过程和活动所表现出的特征通常都采取较理想化的处理手段, 即没有充分考虑“产品设计活动”所表现出的复杂性和不确定性。这主要是由于不确定因素的难以把握和度量造成的。因此, 难以从真正意义上实现设计过程的优化



分析,进而不能提供一套较好的设计过程规划和设计项目管理的量化模型。采用“量化模型”的规划方法能分析设计任务中的不确定因素,并以此给出一个基于任务量化模型的设计评审点设置过程。

## 5.1 量化法与设计过程规划

该方法是从项目规划和管理的角度,以规划设计流程中各个工作对可重构设计过程进行优化。其中,任务影响因子模型及算法的引入,是为了从任务关键度(critical degree)、任务出错度(error-occurrence degree)以及任务波及度(spread degree)三方面描述设计活动中任务对整个项目的直接和间接的影响程度。在此基础上,在关键任务后设置评审点是为了从设计流程上保证质量,及早发现并解决设计中出现的问题,避免由于关键任务的失败而导致项目的大循环。评审规划是设计过程规划的重要内容之一。

### 5.1.1 设计评审规划技术及其研究

设计评审(design review)即是对一个设计项目中已经进行的工作部分进行审核和检查,以期及时发现设计中的错误和缺陷并及时改正,从而保证整个项目的顺利进行。近年来,许多研究者认为从研究设计评审技术的角度来提高设计过程,性能十分重要,并有研究人员针对设计评审相关技术展开了多方面研究。Fu, M. C. 和 East, E. W. 引入了一个虚拟现实设计评审系统来允许多个设计人员同时在一个未结束的建筑模型的三维表示界面上工作,来共同评审已经完成的部分工作以确保整个项目的顺利进行。Bochenek Grace 和 Ragusa James 采用一种合成技术来进行产品概念设计的评审工作,他们还在产品设计评审中集成 3D 引导系统使得设计评审超越常规的经验测试。此外, J. S. Busby 从设计人员与其所处环境发生的一系列交互过程的角度来专门研究设计过程中出现的错误,其目的是提高设计过程的性能。Miyake and Fuji 开发了一个基于量化法的软件开发过程的设计评审框架。

### 5.1.2 量化法及相关研究

“量化法(metrics)”是从数量上刻画实体特征的一种方法。其功能可划分为三种:评估、控制、性能测量。

量化法有以下特点:

1) 它是一种更加系统的方法,克服了主观评估带来的偏差。

2) 它是任何过程管理中都十分重要的一种方法。因为“如果一个过程不能被量化,它就不能被管理”。

3) 它是工程领域的一种有效的管理工具。如软件开发设计行业由于使用了量化的方法,因而准确地对工作量、工作难度以及工期进行了预算,并提高了软件产品的质量和生产率。

4) 它是比较、规划以及优化某个过程的基础。

在一些领域,要开发量化法并不是一件很难的事情。然而,在诸如“设计过程”这样高度抽象的领域研究量化法却是一件十分困难的工作。造成这种困难的原因是:设计过程本质上是一种脑力劳动;过程中缺少很容易测量到的因素。尽管存在这样的困难,在软件开发这个特殊的工程领域却已经出现了大量的成果。相对来讲,在其他设计领域,量化法的研究工作则比较有限。尽管如此,从 Johnson 和 Turner 于 1956 年提出的一种“功能转换公式”来量化工程中工作量与销售额的关系。至今,国际上关于“量化法”在产品设计领域的探索已有近五十年的历史。比如,Hamdi A. Bashir 和 Vince Thomson 提出了一个量化的模型来估算设计过程的工作量和时间。Jacome MF 和 Lapinskii, V 提出了一个面向过程的模型来估算电子产品设计项目的工作量。量化法用于决策系统始于 1972 年, Satty 在其研究中采用量化的思想解决了有限资源的规划/分配问题,并提出了其著名的层次分析法原理(AHP)。目前在产品设计领域,仍缺乏一个明确的量化方法用于设计资源规划(如评审规划)。而这种量化的方法对设计项目规划人员是十分有用的,它将有助于对一个即将展开的项目在现有资源环境下做出较合理的规划。

## 5.2 基于量化法的任务影响因子模型

### 5.2.1 模型原理

任务影响因子(Task Influence Factor, TIF)模型的原理如下:不同任务对整个项目的重要程度通过三方面来量化,分别是任务关键度、任务出错度和任务波及度。首先,通过采用层次分析法原理来度量项目中不同任务对整个项目的重要程度,其中允许考虑多准则和用户自定义准则。其次,通过基于实例的推理来分析不同任务的出错倾向程度,新技术的采用、缺乏相关经验与历史出错频度被看作探测一项任务出错的信号。最后,通过任务依赖矩阵计算任务出错的波及程度。三个指标的乘积构成了任务影响因子,它反映了在有限的项目资源下一项任务应受到重视的优先程度。

### 5.2.2 模型计算步骤

#### (1) 确定任务分解及其之间关系

为了扩展所提出方法的适用范围,我们将问题一

般化,即从探讨一般的设计项目规划入手。

一个设计项目(design project)往往可以划分为几个子项目(sub-project),子项目又可以进一步划分为若干任务(task)。一个设计项目或是一个子项目都可以看作是由若干设计任务组成的。用任务分解树表示任务的分解,如图 43.4-8 所示。

**定义 4.1 子项目任务集:**子项目任务集表示一个设计子项目所含的任务单元的集合,用  $S_i$  表示: $S_i = \{T_{ij}\}$ ,其中  $i=1, \dots, p$ ,  $p$  是子项目的个数; $j=1, \dots, m$ ,  $m$  是子项目  $i$  所包含任务单元的个数。如图 43.4-8 所示, $S_1 = \{T_{11}, T_{12}, T_{13}\}$ ,  $S_2 = \{T_{21}, T_{22}, T_{23}, T_{24}\}$  等。

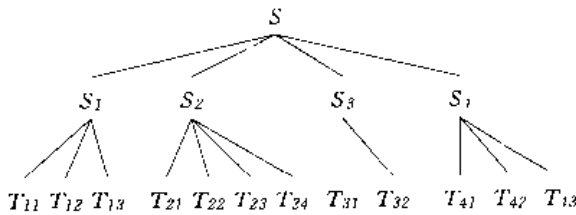


图 43.4-8 任务分解树

**定义 4.2 项目任务集:**项目任务集表示一个设计项目所含的任务单元的集合,用  $S$  表示: $S = S_1 \cup S_2 \cup \dots \cup S_p = \{T_{ij}\}$ 。

如图 43.4-8 所示, $S = \{T_{11}, T_{12}, T_{13}, T_{21}, T_{22}, T_{23}, T_{24}, T_{31}, T_{32}, T_{41}, T_{42}, T_{43}\}$ 。

任务分解是将一个抽象的任务细化为更加具体的任务的过程。经过逐步任务分解,最终得到了一组相对简单和具体、易于处理的任务单元,将这组单元称为一个设计项目的任务集。任务分解及其间关系的确定是一个复杂的决策问题,它是设计过程规划的基础。任务分解的难度在于如何确定任务单元的“粒度”大小。一方面,过粗的“粒度”往往从表面上简化了整个设计项目,却难以暴露那些需要引起重视的任务单元或环节,从而失去了分解的意义;另一方面,过细的“粒度”又削弱了部分任务单元的整体性,导致了任务间衔接工作的增多,并因此带来由于信息传递和人员交接造成的出错率的增加,此外过细的分解也将带来整个系统监控成本的增加,同样不利于设计过程的优化。因此,任务分解的合理程度将依赖于对任务的目标(质量、工期、成本)、资源的了解程度;而设计任务的分解可以采取分层次的解决方案,即不同的分解层面上根据需求采取不同的任务分解“粒度”。比如,设计一台“组合机床”时,项目总规划者只要将该设计任务分解为“刀具及装夹系统”、“定位及进给系统”、“多轴箱系统”、“液压系统”、“电气系统”等不同的设计子项目,而不必了解其中某个齿轮设计任务的具体划分。而某个设计小组也许采取较细的粒度来分解交给他们的任务。多轴箱设计任务将细分为:齿轮设计任务、箱体设计任务、

轴系设计任务等。其中某个人员可能采取更细的粒度来将交给他的齿轮设计任务进一步细分为圆柱齿轮设计任务、圆锥齿轮设计任务等。对于某一层上,那些信息依赖较多的“强耦合”的任务也可以整合为一个任务,即分解的过程可能伴随着局部的整合过程;从项目规划与管理的观点,任务分解应综合考虑经济性、系统可监控性、资源的类型和数量等现状。即任务分解应该能够使得管理者(或管理程序)及时了解各个任务的开始、执行、完成的状态,并在此基础上对任务进行控制与调整,这种监控和调整的成本也应当控制在一定的范围之内。总之,任务分解是一个在多层次上展开的复杂的过程。一个好的任务分解结果本身是由多方面因素决定的,其中也包括规划者的经验。

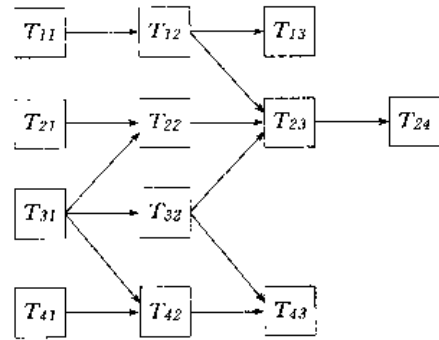


图 43.4-9 任务次序图 G

任务间关系的确定,取决于任务间的次序约束和潜在的技术约束。这些约束是由于各个任务间的技术信息的依赖性(technology dependence)或设计资源依赖性(resource dependence)以及其他因素决定的。任务分解树可以描述任务间的隶属关系和任务逐步细化的过程,却不能反映任务在执行过程中,它们之间的先后次序约束关系。为了描述任务集内的任务在执行中的次序和相互关系,并结合并行开发活动中任务间的复杂依赖关系,采用“任务次序图”来表示任务间的次序约束。如图 43.4-9 所示来表示并行任务的执行次序。当项目规划展开时,图中每个任务都有一个工作量的粗略估计,用  $w_i$  表示。从图中可以看出,并行设计任务展开时,任务间的联系是复杂的。次序图中某些任务的开始执行必须依赖于一个或几个任务的完成。如任务  $T_{23}$  的开展直接依赖于任务  $T_{12}, T_{22}, T_{32}$  的完成。而某些任务的完成情况可能会直接或间接地影响到其后的许多任务。如  $T_{31}$  除对  $T_{32}$  有影响外,还对  $T_{22}$  和  $T_{42}$  有直接影响,对  $T_{23}, T_{24}$  和  $T_{43}$  有间接影响。

## (2) 量化任务关键度

为了分析一个任务的关键程度,需要综合考虑两方面的因素。一方面,对一个给定项目而言,存在不止一个准则来判断其关键程度,如性能可靠性、材料和制造成本、外型美观性等。对不同的产品类型,给定准则

的重要程度是不同的。例如,对一个军用液体发射火炮的设计项目,可靠性的重要性十分突出,外形的美观则相对次要得多。而对一个装饰灯的设计项目,外形美观要求却相当重要。另一方面,给定任务对不同准则的“贡献”也是不尽相同的。例如,O型圈设计任务对液体发射药火炮的可靠性的贡献大于对外型美观程度的贡献。我们用一个圆来表示某评价准则在总目标中的比率(如图43.4-10所示),上半圆中的内容表示该准则的标志,左下1/4圆的内容表示该准则对其上一个节点所代表的准则所占的比率,右下1/4圆的内容表示该准则对总目标所占的比率。

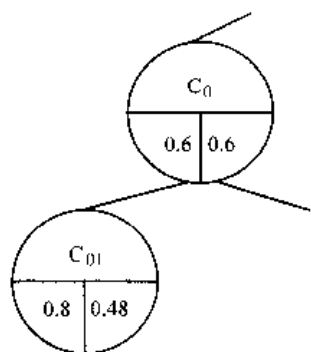


图 43.4-10 权重表示举例

用层次分析法(AHP)解决此问题应分为四步。第一步,建立整个项目的层次分解结构;第二步,建立准则两两比较矩阵;第三步,在给定准则下计算每个准则的相对权重  $m_i$ ;第四步,计算每个任务的组合权重  $C_j$ 。

本模型引入任务关键系数矩阵  $F_{q \times n}$  来表示在每一准则下,任一任务对项目的关键系数,如式 43.4-1 所示。其中,  $q$  表示准则数目,  $n$  表示任务数目。假设给定准则  $R_i$  在总目标中的相对权重是  $m_i$ ;在准则  $R_i$  下,任务  $i$  对项目  $P$  的关键系数为  $F_{ij}$ ;在总目标下,任务  $i$  对项目  $P$  的关键度为  $C_j$ ;则

$$F_{q \times n} = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1j} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2j} & \dots & F_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ F_{q1} & F_{q2} & \dots & F_{qj} & \dots & F_{qn} \end{bmatrix} \quad (43.4-1)$$

对于任意  $i(i=1,2,\dots,q)$ ,存在  $F_{i1}+F_{i2}+\dots+F_{in} = 1$ 。即在某个给定规则  $R_i$  下,所有任务对项目  $P$  的关键系数之和为 1。

各任务在总目标中对项目  $P$  的关键程度  $[c]$  由式 43.4-2 计算而得。

$$[c_1 \ c_2 \ \dots \ c_j \ \dots \ c_n] =$$

$$[m_1 \ m_2 \ \dots \ m_i \ \dots \ m_q] \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1j} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2j} & \dots & F_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ F_{q1} & F_{q2} & \dots & F_{qj} & \dots & F_{qn} \end{bmatrix} \quad (43.4-2)$$

记为:

$$[c] = [m][F] \quad (43.4-3)$$

(3) 量化任务出错度

设计决策人员通过分析以前的案例来做决策,这个方法在分析任务出错时也可采用。通过分析历史设计案例中的错误出现情况,发现三种情况会暗示高的设计错误的出现。第一,设计中采用新技术;第二,缺乏相关的设计经验(尽管没有采用新技术);第三,当企业历史记录表明,该类设计问题在一次迭代后有出错率较高的情况时,采用  $\beta^1, \beta^2$  和  $\beta^3$  来分别度量三个因素。进一步分析可以发现,任务关键程度是一个任务间的相对值,而错误出现率的本质却不是由任务间的比较而定,它取决于任务的实际情况。所以仍然采用 AHP 方法是不恰当的,但是可以参考 Satty 在 AHP 方法中提出的系数标定方法(如表 43.4-4 所示)。

表 43.4-4 AHP 中的系数涵义

A/B 的比率	中文涵义(译前涵义)
1	对 C, A 与 B 同等重要
3	对 C, A 比 B 重要一点 (little more important than)
5	对 C, A 比 B 重要相当多 (much more important than)
7	对 C, A 比 B 绝对重要 (severe more important than)
9	对 C, A 比 B 极端重要 (severe more important than)
2, 4, 6, 8	在 1 与 3, 3 与 5, 5 与 7 和 7 与 9 之间表示折衷的中间值

引入 5 个系数(4,3,2,1,0)来标定每一种情况的程度。以一项新技术的掌握程度从低到高标定  $\beta^1$  为 4/3/2/1/0,  $\beta^2$  取值 4/3/2/1/0 分别表示所面临的任务的经验从少到多(具体界定标准由企业规划人员裁定,例如,4/3/2/1/0 表示该类任务的经验次数 0/1 次/2 到 5 次/5 到 20 次/20 次以上),当  $\beta^2$  的取值非 4 时,即以前曾经开展过类似任务时,  $\beta^3$  取 4/3/2/1/0 分别表示该任务的第一次迭代中出现错误的百分率为 90%~100%/60%~90%/20%~60%/0%~20%/0。并且特别规定当  $\beta^2$  为 4 时,  $\beta^3$  取 4。一个任务的错误出现率( $E_i$ )由  $\beta^1, \beta^2, \beta^3$  之和决定(如式 43.4-4 所示):

$$E_i = (\beta^1 + \beta^2 + \beta^3)(i = 1, 2, \dots, n) \quad (43.4-4)$$

#### (4) 量化任务波及度

任务间的信息或资源的依赖关系造成了由于某个任务的延误而给其他任务带来的延迟。任务波及度用来度量由于一个任务的延迟带给整个项目的总延迟。引入任务依赖矩阵(task dependence matrix)来量化任务波及度。当任务分解完毕后,可以开始分析任务间的依赖关系,这种依赖取决于信息的依赖和资源的限制。任务次序图反映了任务间的依赖关系。值得注意的是,任务次序图中的箭线约束关系并不直接等于任务依赖矩阵表达的依赖关系,通常前者是后者的真子集。例如:在任务次序图中, $T_{31}$ 的后继任务为 $T_{22}$ 、 $T_{32}$ 与 $T_{42}$ ,而任务依赖矩阵中可能存在 $T_{23}$ 、 $T_{21}$ 与 $T_{43}$ 也将依赖 $T_{31}$ 。这些任务对 $T_{31}$ 的依赖程度可能存在不同,在设计规划阶段,缺乏进一步的量化细节,因此假设 $T_{31}$ 的延迟会给这些任务的决策带来反复(iteration)。在任务依赖矩阵 $A$ 中, $a_{ij}=1/0$ 分别表示任务 $j$ 对任务 $i$ 存在/不存在依赖。设已知任务 $j$ 的工作量为 $w_j(j=1, 2, \dots, n)$ ,则任意任务 $T_i$ 的波及度(如式 43.4-5 所示):

$$T_i = \sum a_{ij} w_j \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, n). \quad (43.4-5)$$

#### (5) 计算任务影响因子

当一个任务的上述三项(任务关键度、任务出错度和任务波及度)都量化完成后,其任务影响因子可以通过式 43.4-6 的计算得到:

$$D_i = C_i \cdot E_i \cdot S_i = C_i \cdot (\beta^1 + \beta^2 + \beta^3) \cdot S_i \quad (i = 1, 2, \dots, n) \quad (43.4-6)$$

### 5.2.3 几点说明

1) 量化法在设计评审规划中的应用有助于提高设计过程的定量化。

2) 该任务影响因子模型可以给有限资源下的规划活动提供基础和参考。

3) 同其他很多量化法一样,量化的过程不可避免地存在一定程度的主观因素,但这并不影响通过不断完善量化指标来对其进行使用。实际应用中,通过具有开放结构的企业数据库的不断修正和完善,可得到该模型中量化系数的合理取值。

## 5.3 基于任务影响因子模型的设计评审规划

### 5.3.1 需求量化的基本思想

当一个项目在有限资源的环境下展开时,项目规划人员希望通过其规划过程将资源进行最合理的分

配,即将有限资源分配给需求较迫切的对象。我们从项目的角度出发,将“设计评审”看作是项目能够提供的一种资源。设计评审作为一类活动,本身需要耗费人力、财力以及时间。根据任务影响模型,我们将任务对项目的“关键程度”的大小看作是任务对评审资源的“需求程度”的一种量化。在资源有限的情况下,重要的任务、易于出错的任务、出错后对整个项目造成的影响较大的任务应当得到优先的评审。因此,根据 $D_i$ 的大小次序和现有评审资源的数量可以为任务进行评审点设置,即将现有的评审资源设置到相对较重要的若干个任务之后,来提高评审效率。

### 5.3.2 步骤

基于任务影响因子模型的设计评审规划技术分为两个步骤:首先根据任务影响因子模型计算项目中各个任务的影响因子 $D_i$ ;其次确定项目所能提供的评审资源的个数 $N$ ,并将评审点设置到按照 $D_i$ 的大小排序的前 $N$ 个任务后。

### 5.3.3 应用举例

该应用项目是一个摩托车新品设计项目。设计过程可以分解为以下 10 个任务如图 43.4-11 所示:①总体设计、②鞍座设计、③灯具设计、④前护板设计、⑤车架设计、⑥发动机设计、⑦构件选型设计、⑧覆盖件设计、⑨构件选型确认、⑩样车试制。

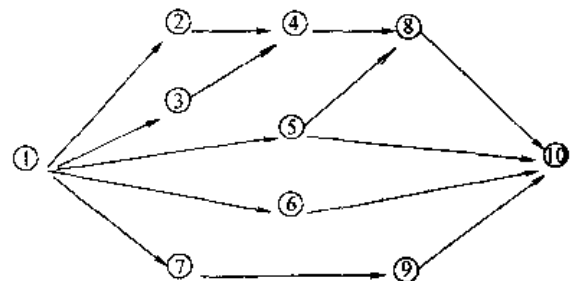


图 43.4-11 任务流程图

由上图可知:

1) 该设计过程中包含构件选择部分(如任务⑦、⑨)和新构件开发/改进过程(如任务②、③、④、⑤、⑧、⑥)部分,是一个具有可重构设计过程基本特征的实例。

2) 严格地说,任务⑩并不属于设计阶段,因而后续分析过程将不涵盖其相关量化过程。图中包括该任务是为了明确给出各任务的流向节点。

3) 我们中的“评审”指项目内部规划展开的设计审核、检查,类似于企业常说的“内部评审”的概念。

实例分析的具体过程:

(1) 给出任务次序图

如图 43.4-11 所示。任务流程图中将可以并行展开的任务(不相关或弱相关任务)尽量并行展开;对存在次序约束的任务,用箭线表示其间约束。

(2) 量化任务关键度

同一任务对不同的评价准则会有不同的关键程度,在该项目规划中,我们取“可靠性”,“经济性”和“外

观”,作为三个评价准则。分别反映了从行业规范和项目目标角度分别对车体性能(可靠性、装配无干涉等)提出的要求。

采用层次分析(如图 43.4-12 所示)确定各任务的关键系数  $F_{ij}$ (结果如图 43.4-13 所示)。

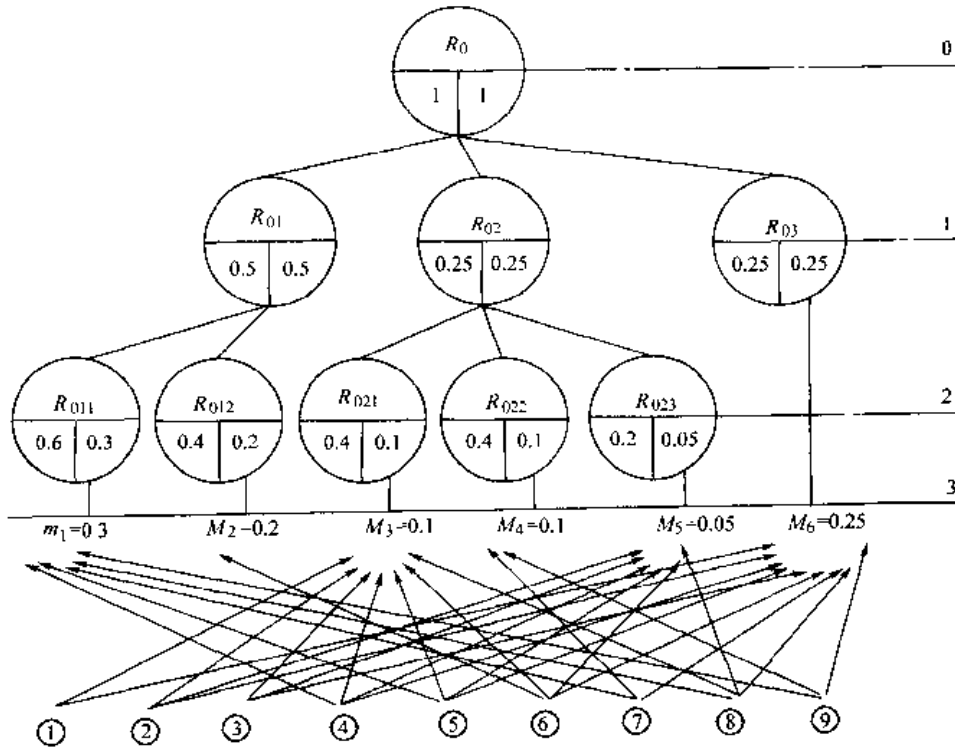


图 43.4-12 任务关键度 C 的层次分析过程

$R_0$ —项目总指标  $R_{01}$ —可靠性  $R_{011}$ —无断裂/干涉  $R_{012}$ —无泄漏  $R_{02}$ —经济性指标  $R_{021}$ —低制造及材料成本  $R_{022}$ —低外包成本  $R_{023}$ —低维修成本  $R_{03}$ —美观/满足个性化需求

各任务的关键度为:

$$C_i = \sum m_j F_{ij} = m_1 F_{i1} + m_2 F_{i2} + m_3 F_{i3} + m_4 F_{i4} + m_5 F_{i5} + m_6 F_{i6} \quad (i=1, 2, \dots, 10)$$

在该例中,  $C_1=0.065, C_2=0.0475, C_3=0.1025, C_4=0.105, C_5=0.10, C_6=0.2475, C_7=0.1675, C_8=0.10$  以及  $C_9=0.065$ 。

(3) 量化任务出错度

任务出错程度的量化采用前述的三个指标  $\beta^1, \beta^2, \beta^3$ 。 $w_i$  表示了各个任务的估计工作量(work amount/work load)。表 43.4-5 给出了相关计算结果。

表 43.4-5 各任务的  $w_i$  和  $\beta^1 + \beta^2 + \beta^3$

	1	2	3	4
$w_i$	15	10	8	5
$\beta^1 + \beta^2 + \beta^3$	0+0+2	0+2+2	0+4+2	0+0+2
	5	6	7	8
$w_i$	8	40	30	10
$\beta^1 + \beta^2 + \beta^3$	0+0-2	0+0-4	0+0+3	0+0+2
				0+0-1

(4) 量化任务波及度

先确定任务间的依赖关系(如图 43.4-14 所示),然后按照前述方法计算任务波及度。 $S_1=w_1+w_2+w_3+w_4+w_5+w_6+w_7+w_8, S_2=w_2+w_4+w_8, \dots$ , 和  $S_9=w_9$ 。

$$S_1 = 141; S_2 = 25; S_3 = 23; S_4 = 15;$$

$$S_5 = 18; S_6 = 50; S_7 = 45; S_8 = 10; S_9 = 15。$$

(5) 计算任务影响因子并设置评审点

$$\text{根据 } D_i = C_i \cdot E_i \cdot S_i = C_i \cdot (\beta^1 + \beta^2 + \beta^3) \cdot S_i$$

$$D_1 = 18.33; D_2 = 4.75; D_3 = 14.145; D_4 = 3.15;$$

$$D_5 = 3.6; D_6 = 49.5; D_7 = 22.6125; D_8 = 2;$$

Task	1	2	3	4	5	6	7	8	9
$m_1$	—	—	—	0.2	0.2	—	0.3	0.2	0.1
$m_2$	—	—	—	—	—	1.0	—	—	—
$m_3$	0.15	0.2	0.2	0.1	0.1	0.15	—	0.1	—
$m_4$	—	—	—	—	—	—	0.7	—	0.3
$m_5$	—	0.05	0.15	0.2	0.1	0.4	—	0.1	—
$m_6$	0.2	0.1	0.3	0.1	0.1	0.05	0.03	0.1	0.02

图 43.4-13 任务关键系数

$F_{6 \times 9} =$ 

$$\begin{bmatrix} 0 & 0 & 0 & 0.2 & 0.2 & 0 & 0.3 & 0.2 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 \\ 0.15 & 0.2 & 0.2 & 0.1 & 0.1 & 0.15 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0.3 \\ 0 & 0.05 & 0.15 & 0.2 & 0.1 & 0.4 & 0 & 0.1 & 0 \\ 0.2 & 0.1 & 0.3 & 0.1 & 0.1 & 0.05 & 0.03 & 0.1 & 0.02 \end{bmatrix}$$
 $D_9 = 0.975。$ 

按  $D_i$  的大小依次是  $D_6, D_7, D_1, D_3, D_2, D_4, D_8, D_5$  和  $D_9$ 。

假设仅能提供 4 个评审点, 则应将其置于  $D_6, D_7, D_1$  和  $D_3$  之后。

Task	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	—	1	—	1	—	—	—	1	—
3	—	—	1	1	—	—	—	1	—
4	—	—	—	1	—	—	—	1	—
5	—	—	—	—	1	—	—	1	—
6	—	—	—	—	—	1	—	1	—
7	—	—	—	—	—	—	1	—	1
8	—	—	—	—	—	—	—	1	—
9	—	—	—	—	—	—	—	—	1

图 13.4-14 任务依赖矩阵 A

#### (6) 结果比较

通过采用文中提出的基于量化原理的评审规划方法, 在项目实例中, 从项目资源规划的角度, 任务对项目的重要程度依次为⑥发动机设计、⑦构件选型设计、①总体设计、③灯具设计、②鞍座设计、①前护板设计、⑤车架设计、⑧覆盖件设计和⑨构件选型确认设计。当项目资源的数量有限时, 上述各任务在享有资源的优先程度上可以参照该顺序。

以上结果是采用“量化法”进行任务分析所得结果。下面从实际项目执行历史记录中, 查找出与评审相关的实际情况(具体表现为设计出错/缺陷实例), 用以对比。

在实际项目开发过程中, 经过阶段审核(企业内部评审)和行业审核, 以及成品市场反馈, 共出现以下设计出错/缺陷: ①发动机的设计中出现“油液泄漏”情况。原因是由于密封结构及其相配合的其他结构的尺寸不合理。②总体设计中出现概念设计缺陷: 鞍座造型过于狭小, 不够舒适, 且与该车型的整车重量不协调, 市场接受程度差。原因是总体设计阶段市场调查不够。

③灯具设计中, 头灯造型不合理。头灯模具尺寸不合理。④前挡板轻微变形, 原因是选材的失误。由于该项目属于车型外观改进类项目, 以上设计失误中, ①、③情况的出现, 导致“样车试制”的拖延, 进而而在原定的产品市场交付时间基础上延误了 4 个月, 给本项目带来了很大的“机会成本”损失。鞍座的造型设计缺陷直到产品后期交付市场才发现, 影响了产品的市场销售, 并对企业形象造成一定的损失。事后的分析发现, 其实这些错误/缺陷的出现并非偶然, 而是能够从企业的错误案例统计中暗示的, 然而, 由于缺乏量化的分析手段, 仅仅凭借感觉难以把握错误的出现阶段及其波及程度, 从而难以作出合理的资源规划。

作为对比, 应用我们提出的基于量化分析法的评审资源规划后, 以上的各任务都被列为具有较高资源占有优先度的任务, 当能够提供四个评审点资源的情况下, 上述出错任务中的三个得到了评审优先。这个项目实例仅是一个较局部的例子, 但是它能够说明量化分析在包括评审规划在内的项目资源规划中的用途和相应的作用。值得说明的是, 合理的规划不仅取决于一个较全面的任务量化分析工具, 它还取决于企业自身特点、以及管理者的经验和智慧。本方法从任务分析的角度, 为资源的规划提供了一个初步的量化工具。

从设计环节保证产品质量是设计活动的目标之一。为了保证质量, 经常需要在设计流程中设置一些评审环节, 经常地、有效地评审活动是有意义的, 它是确保“及早发现问题”的有效手段。另一方面, 从项目规划的角度来看, 评审环节作为一个活动, 它的增设又会给项目带来消耗成本、延长工期、需要资源等负面的影响。那么如何寻找问题的最佳折衷点呢? 传统的设计中, 由于对产品设计活动从项目规划角度来进行计划的做法并未推广, 因此, 评审点的定位是根据直觉的, 带有很大的随意性。在认为有可能重要的任务后设置评审环节, 或者不惜成本, 大范围设置评审环节。随着设计方法的发展和多学科的交叉, 产品设计过程正向着逻辑化、层次化、定量化发展。上面的基于任务影响因子的任务模型为评审点设置的优化提供了基础, 从而可以避免仅凭借直觉、完全依赖主观的方法所造成的弊端。我们提出的任务影响因子(TIF)模型不仅可以应用于可重构设计过程规划中, 亦可为一般设计项目中的类似资源规划活动提供基础和参考。

## 第5章 模块化技术的应用

### 1 概述

自20世纪60年代以来,模块化从概念走向应用,在近20年来模块化技术在理论上更加完善,并得到了进一步的发展。在学科交叉的影响下,模块化技术与并行工程的结合,模块化技术与集成技术的结合,使模块化技术得到更广泛的应用。从设计到制造,从加工到管理,从电子产品到工业产品,从工业生产到教育事业,到处可见模块化技术的应用。

随着电子技术的发展,电子器件经历了从分立器件到集成电路的变革,而集成电路的由小规模发展到中规模直到大规模集成电路的出现,均体现了变革的深入性。由于受到各种条件的限制,如材料、加工方法及投资等,集成电路的缺点就显现出来。

20世纪80年代以来,模块化的微电子器件应运而生,并大大促进了计算机的发展,计算机存储量的不断加大,速度的不断加快,功能的不断扩展均因得益于模块化电子器件的应用。

计算机的程序设计则是应用模块化技术的典型范例之一。

一个较大型的CAD程序的设计将其各部分的功能划分成许多相对独立的模块,每个模块都有特定的功能,有自己的信息入口和信息出口,可以在不必了解模块内部结构时,就能方便地将它们组合起来形成一个较大的应用程序。设计模块化的程序应注意下述几点:

- 1) 各程序功能模块的联接与覆盖。
- 2) 各程序功能模块间的数据传递。
- 3) 各程序功能模块的运行管理。

20世纪80年代以来,模块机器人概念及技术的研究一直是机器人领域中热门的课题之一,由于机器人在设计时具有强烈的针对性,即完成什么样的动作,采用什么样的结构等,而机械结构限制了机器人适应不断变化的作业要求的能力,因此人们就非常需要通过自身的结构型式的改变来完成无法预知的作业任务。可重构机器人也就成为研究的重要方面,而这种可重构性也正是运用模块化技术经不同组合而重构为满足多方面需要的机器。按机器人的功能要求大致可划分为关节模块和连杆模块两大部分,但关节模块又是基本模块。这一模块与其他机械方面构成的模块的区别就在于其是一个一体化的基本单元,集通讯、控制、

驱动和传动一体化,该模块的本身就是—个集成系统。1988年美国卡纳基梅隆大学研制出了一种可重构的模块化机械系统。这可以看作是机器人领域运用可重构理论的最早的样机之一,目前国外已经在开发动态可重构机器人系统和静态可重构机器人系统两大类,这些均是在模块化技术的基础的扩大应用。

模具也是模块化技术应用的典型,从20世纪80年代初,就已完善冷冲模、橡胶模、塑料模的典型零件的标准化,在此基础上又完善了各类模具的模架系统,通过组合拼装技术把定位、紧固、导向等多种标准化元件加上型芯组合就可形成新的模具。

随着模块化技术应用范围的扩展,模块化产品(系统)的特性也逐渐的显露出来。

模块产品(系统)具有如下的效应:

#### (1) 重构效应

通过由一组具有不同结构、不同尺寸或不同功能的可互换的模块进行组合,而重新构造新产品。该重构效应大大提高了产品的柔性。

#### (2) 功能效应

模块是功能的载体,模块或者组合后的高一级的模块均是功能的具体体现。

#### (3) 互换效应

模块之间的互换性,将是系统与系统之间的互换,这不仅体现在接口界面的互换,这集中体现了系统与系统之间的性能参数的匹配互换。

#### (4) 扩展效应

模块化产品的设计尽量以某一型号为基型,但设计出来的产品是成组的,这种设计完全是不断补充的,通过不同模块的重组又能得到成组的新产品,产品结构可得到不断地扩展。

#### (5) 开放效应

由于现代社会中技术发展的飞跃,新技术从实验室到应用的时间越来越短,因此利用新技术开发的模块将不断地补充到模块库中,使产品更新换代的周期大大缩短,这种模块化产品的开放效应为企业争取了时间,为市场竞争提供了有力支持。

#### (6) 适应效应

根据不同的需要,在很短时间内构筑了新产品以适应市场的需求,响应的速率则要取决于模块库的结构和开发,模块库的建立要有一种快速自适应能力。

## 2 模块化在造船业中的应用

国外早在 20 世纪 70 年代,模块化技术在造船业中已得到了应用,西方某船厂在军舰上使用了三十多个模块,如 MEICO(多用途综合)的舰艇,开发了模块化的武器系统和电子系统。功能单元系统(FES)就意味着将组成一个完整的武器系统或电子系统的全部子系统和附属单元集成一个易于互换的单元模块。它可以安装于集装箱、托盘或构架。它们的实践证明,800~4000t 舰艇所用的对空/对地火炮、对空导弹系统、对地面导弹系统、对潜式武器火箭发射架等及北约组织的现有武器除海标枪导弹和标准导弹外,都可以装在底面构架上构成功能模块。又如丹麦的舱室模块是总体吊装的。日本在 20 世纪 80 年代也采用模块造船,国外的很多工厂生产的均是模块商品如主辅机供油模块、分油机模块等。

尽管国内的模块化造船在 20 世纪 90 年代已研制成功了凸轮滑油模块并投入生产,但真正开展这方面的研究始于 1991 年,在船舶总公司召开的 35000t 级浅水散货船综合标准化工作会议上予以确定,明确通过两年的研制完成 13 个模块课题的任务。1991~1993 年所完成的 9 个模块的设计、上船试验及应用取得了较大的成果,同时也显示了模块化造船的优越性。

模块化造船显示出了如下的优点:

1) 缩短设计周期。我国是世界上的造船第三大国,但效率不高,利润低,其重要原因在于设计周期和船台周期太长。

在造船系统中 20 世纪 90 年代

设计周期一般 9 个月

制造船台周期 3 个月

下水试航 2 个月

共计 14 个月,而设计周期占整个周期的 64% 左右。

当 38800t 货船我国要用 60~80 万工时,而日本仅为 20 万工时,就相差了 3~4 倍。当采用模块化造船后,船台周期从原 90 天缩短到 40 天,船台码头一天就几十万元。我国一个船台一年可以下 4 条船左右,但日本大阪船厂一个船台一年要下 6~9 条船,我国仅是日本的 2/3~1/2。日本造船所以有这么短的周期主要是采用模块化造船。国外的初步统计表明,当采用模块化后,船舶设计周期一般可缩短 2/3~1/2。

2) 提高效率,加快制造周期缩短工时,可大大降低成本,便于专业化生产。由于模块的通用性,使原来单件生产的产品利用其共性变成批量生产,把模块商品化,使造价降低了 1/3。

3) 打破过去船厂仅以船作为产品的格局,船厂可

以用各种船用模块进入市场,作为商品生产(即中间产品),由此可使船厂产生经济效益。

4) 维修方便,由于模块划分,功能分清,在故障诊断中非常容易发现问题,布置也有规律性。

5) 提供了良好的人—机—环境的条件,提高工作效率,有的模块完全可以在陆上试验、调试、安装。

6) 大大提高船舶可靠性,其中每一模块都进行例行试验,在安装时只要连接即可,舾装的安全性得到了有效的提高。

7) 为造船系统的 CAD 提供了必要的基础。

模块一定要系列化,要搞菜单,即设计目录。

在造船工业中,过去也曾采用过单元组装这种方式进行造船,但单元组装与模块不能混为一谈,因单元组装是根据其工艺性、安装性,应用组合化原理而引出的概念,组装单元可能不是一个完整模块,而是某一模块中的一部分,也可能是几个完整模块的组合体。而在目前造船中采用的分段建造法则结合了船舶的特点,运用模块化思想划分的一种有效方法,即模块的划分以区域和功能相结合为原则,如图 43.5-1 所示。



图 43.5-1 船体模块的区域划分

1—船艏功能模块 2—过渡总段 3—平行舢体

4—船艉功能模块 5—上层建筑功能模块

由图可知:船体可以由船艏、船艉、上层建筑等功能模块和若干平行舢体结构模块,再加上专用过渡总段组成一艘模块化的船舶,如改变图中的过渡总段和平行舢体结构模块,即可改变船舶总长度,排水量组成不同用途的船舶,如散装货轮、油轮、矿砂轮等,并使船舶性能保持在合理范围内。

船员舱室是船员日常生活的住室,它比一般的寝室住的时间还要长,因此如何为船员营造一个舒适的环境是非常重要的,但因船员的等级和工作需要其功能也是不一样的。丹麦的船员舱室的设计是世界上著名的,它们可以按不同需要进行不同船员舱室快速、灵活的整体装配,这主要是根据该模块的功能分解图,通过标准化而达到的,图 43.5-2 是船员舱室的功能分解图:

三种情况可用不同的编码来体现其功能需求,只要把标准化对象选为舱室面积大小,这样就明显区别出级别,但不要去把室内的布置作为标准化对象,否则将不利于舱室的制造。



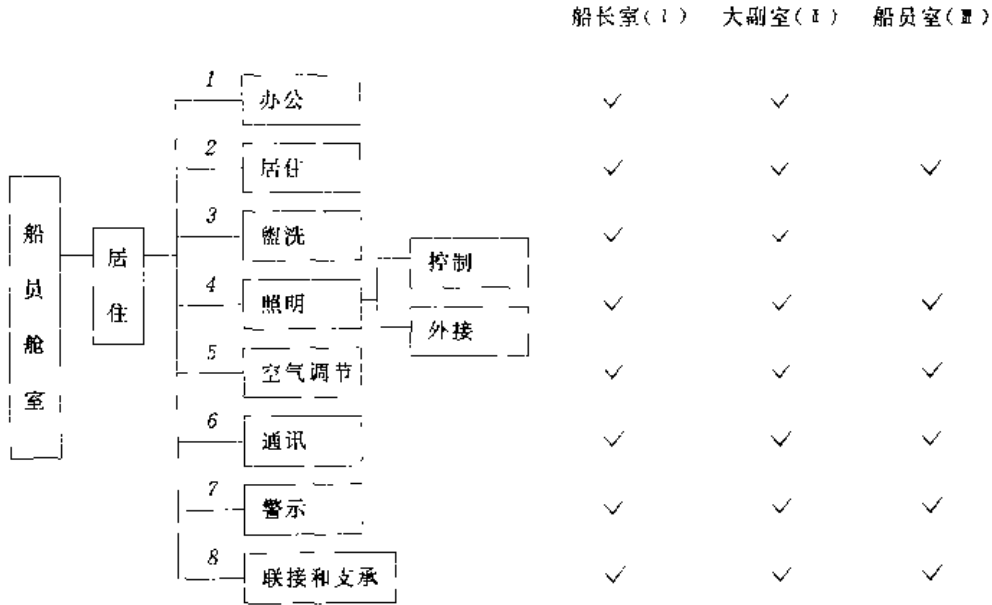


图 43.5-2 船员舱室功能分解图

### 3 模块化在临床医学中应用

口腔医学在临床医学中占有相当重要的地位,而牙齿又是口腔医学的重要方面,至今人们对牙齿的要求也逐渐在提高,除了完成咬嚼作用以外,又加上美观,因此补牙、装牙人流不息。目前在装牙过程中用印膜的办法制作基本上是单件制作。人们就想能否建立一个库,装牙时从库中选取,届时只要选取匹配的颜色即可。该库的建立则依据牙颌组织的解剖学特征和生物力学特征。这就是几何与力学仿真模型的建立,为此通常是以给出的基于 DICOM 格式的 CT 数据处理详细过程,并进行试验,再用层切法反求完整的牙列咬和面,其精度达到 0.2mm,然后在统一坐标下,对数据配准,经三维重建后,

完成牙颌组织几何学仿真建模并进行力学仿真建模。在建立的牙颌组织几何与力学仿真模型库中针对牙齿特点把模型划分为模块,使用时重新调配、组装满足需要,其中应用模块化思想快速高效地建立了大量牙颌组织正常、病理和修复状态的有限元模型。运用该模型能对口腔中各种修复体(修复体指的是人工器官、人工牙、义齿、人工下颌关节等)及基牙牙周组织进行生物力学研究,由此了解各种修复体在不同载荷方向,不同加载部位,修复体及基牙牙周组织的不同应力分布的情况下,为嵌体、冠、桥及可摘局部义齿、全口义齿、颌面修复等的设计与优化以及基牙的选用提供计算依据。图 43.5-3 所示为牙颌组织模型库结构。

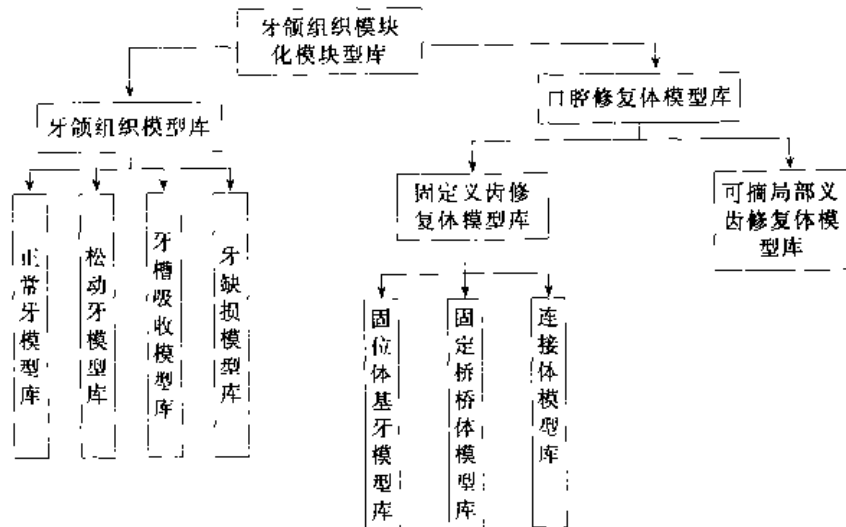


图 43.5-3 牙颌组织模型库结构

## 参 考 文 献

- 1 刘之生,黄纯颖主编. 反求工程技术. 北京:机械工业出版社,1992
- 2 王成焘主编. 现代机械设计. 上海:上海科学技术文献出版社,1999
- 3 辛一行主编. 现代机械设备设计手册. 北京:机械工业出版社,1996
- 4 刘恒,虞烈,谢友柏. 现代设计方法新产品开发. 中国机械工程,1999,19(1):81~84
- 5 Peter Dietz. 产品优化和借助于设计系统学的产品开发过程(1). 李贵轩译. 机械设计,1999(2):1~4
- 6 刘溪涓,蒋寿伟. 可重构设计中基于最小损失函数的零件选择算法. 计算机辅助设计与图形学学报,2002,14(10):967~971
- 7 Bashir HA, Thomson V. Metric for design projects;a review. Design Studies,1999,20(3):263~277
- 8 Tavares L V. Advanced models for project management. Boston:Kluwer acadymic publishers, 1999
- 9 Fu M C, East E W. Virtual design review. Computer-Aided Civil and Infrastructure Engineering,1999,14(1):25~35
- 10 Bochenek, Grace M., Ragusa, and James M. Comparative testing of virtual environment display devices for conceptual design reviews. American Society of Mechanical Engineers, Material Handling Division, MHD,1999
- 11 Bochenek, Grace M., Ragusa, and James M. Integrating virtual 3-D display systems into product design reviews: some insights from empirical testing. International Journal of Technology Management,2001,21(3):340~352
- 12 Busby J S. Error and distributed cognition in design. Design Studies,2001,22(3):233~254
- 13 Liu X, Wang Y, Jiang S. A metrics based task analysis model for design review planning. Design Studies(Elsevier),2003,24(4):375~390.
- 14 Liu Xi-Juan, Jiang Shou-Wei. Multistrategies in case based design knowledge management on web Journal of Dong Hua University (English Edition),2003,20(1):94~99

# 第 44 篇 优 化 设 计

主 编 王石刚  
编写人 王石刚 陆金桂

# 第 1 章 优化设计算法原理

## 1 优化设计概念

### 1.1 优化设计基本概念

所谓优化设计就是指在一定约束情况下,即满足各种设计条件时,利用数值优化计算方法得到产品最佳设计值。利用优化设计,可进一步改善和提高产品的性能;在满足各种设计条件下减少产品或工程结构重量,从而节省产品成本消耗、降低工程造价;可以进一步提高产品或工程设计效率。因此,优化设计是直接提高产品设计性能、降低产品成本的有效设计方法。优化设计可给企业带来直接的经济效益,从而提高企业产品的竞争能力。

优化设计的目标是使设计对象最优,而优化设计的手段是计算机及优化计算软件。优化计算软件是以优化计算方法为基础而形成的应用程序系统。因此,优化设计还可以被理解为采用计算程序的从设计空间搜索最佳设计方案的现代设计手段。优化设计与常规设计中优化计算方法的数学基础包括线性规划、非线性规划、动态规划、几何规划等内容的数学规划理论。

优化设计一般包含如下主要内容:①将设计中的实际物理模型抽象为数学模型。确定设计过程中主要的设计目标和设计条件,在此基础上构造评价设计方案的目标函数和约束条件等。②数学模型的求解。根据数学模型的性质,选择合适的优化方法,并利用计算机进行数学模型的求解,得到优化设计方案。

任何机械设计问题,总是要求满足一定的工作条件、载荷和工艺等方面要求,并在强度、刚度、寿命、尺寸范围及其他一些技术要求的限制条件下寻找一组设计参数。因此机械优化设计问题就是在满足一系列设计参数的限制条件情况下优选一组设计参数,使得设计参数对应的设计指标达到最佳值。机械优化设计问题在数学上可以表达为以等式或不等式函数描述的约束条件和以多变量函数描述的优化设计目标,这就是优化设计的数学模型。优化数学模型中包含了目标函数、设计变量、约束条件等。一般称优化设计数学模型中的目标函数、设计变量、约束条件为优化问题的三个要素。

在优化设计数学模型中,用多变量函数描述的优化设计目标用来评价方案优劣的指标,被称为目

标函数。目标函数一般是可变化的设计参量的显函数。

在优化过程中进行调整的各独立设计参数称为设计变量;设计变量可以是几何参数,也可以是物理参数等。根据设计要求需要预先给定的参数,不能作为设计变量,这些设计参数被称为设计常量。优化设计的数学模型中对设计变量进行了一定的限制,这些限制设计变量取值的等式或不等式函数,称为约束条件。

### 1.2 优化数学模型

优化设计数学模型的标准形式如下所示:

$$\left. \begin{array}{l} \min. f(x) \quad x \in R^n \\ \text{s. t. } h_v(x) = 0 \quad v = 1, 2, \dots, p \\ g_u(x) \leq 0 \quad u = 1, 2, \dots, m \end{array} \right\} \quad (44.1-1)$$

式中  $f(x)$ ——目标函数;

$x$ ——设计变量,  $x = [x_1, x_2, \dots, x_n]^T$ ,  $x \in R^n$ ;

$h_v(x)$ ——等式约束函数,  $h_v(x) = 0$ ;

$g_u(x)$ ——不等式约束函数,  $g_u(x) \leq 0$ 。

工程实践中的设计问题可能会要求多个设计指标达到最优,因此优化问题的数学模型有时会有多个目标函数。如果数学模型中目标函数只有一个,这样的优化问题就是单目标优化问题。如果目标函数有一个以上,这样的优化问题就是多目标优化问题。优化模型中目标函数十分关键,因为它直接影响着优化设计的效果。

在优化数学模型中设计变量的数目可称为优化问题的维数。如果优化数学模型有  $n$  个设计变量,则称该优化问题为  $n$  维优化问题。设计变量数目为  $n$  的优化设计问题,其设计空间是由  $n$  个独立变量坐标轴构成的空间,是  $n$  维实空间  $R$ 。优化问题设计变量的数目越多,其设计空间的维数越高,因而在设计空间寻优的难度也就越大。一般说来,优化设计问题求解的计算量是随设计变量数目的增多而显著增加的。

在优化设计数学模型中,有些设计变量只能取离散的值(例如齿轮的齿数等),这样的设计变量称为离散设计变量。设计变量取值没有限制可以取连续量,这样的设计变量称为连续设计变量。

在优化设计数学模型中,如果约束条件是用数学不等式来表示,可将约束条件称为不等式约束条件。除了不等式约束条件,还会遇到用等式来表示的约束条

件,这种约束条件可称为等式约束条件。等式约束条件对设计变量的约束较严格。一个等式的约束条件等价于两个不等式约束条件,即  $g(x) = 0$  等价于  $g(x) \leq 0$  和  $-g(x) \leq 0$ 。

优化设计数学模型的约束条件是由实际的设计要求决定的。一般可将约束条件分为边界约束和性能约束两种。边界约束是指直接限制设计变量取值范围的约束条件,例如

$$\begin{cases} g_1(x) = a_1 - x_1 \leq 0 \\ g_2(x) = x_1 - b_1 \leq 0 \end{cases} \quad (44.1-2)$$

所谓性能约束,就是由必须满足的设计性能要求决定的约束条件,如行星齿轮传动的装配条件、邻接条件、传动条件,曲柄摇杆机构的曲柄存在条件等。优化设计数学模型约束条件越多,优化计算越复杂。

在优化设计问题中由于约束条件的存在,可将优化问题的设计空间分为可行域和不可行域二个区域。所谓可行域设计空间,就是指满足优化模型的约束条件设计空间;同样,不可行域设计空间就是指不满足优化模型约束条件的设计空间。如果优化设计问题的解位于可行域设计空间,那么优化设计问题解就是可行解或可行点。优化设计问题的可行解或可行点可以理解为是满足约束条件的设计解。可行域设计空间是由所有优化问题的可行点组成的集合。同样,如果优化设计问题的解位于非可行域设计空间,那么优化设计问题解就是非可行解或非可行点。优化设计问题的非可行解或非可行点可以理解为是不满足约束条件的设计解。非可行域设计空间是由所有优化问题的非可行点组成的集合。例如设计点  $x_1$ , 如果  $x_1 = [2, 0]^T$  对应的目标函数值  $f(x_1) = 0$ , 但这个点违反了约束条件  $g = x_1^2 - x_2 + 1 \leq 0$ , 故设计点  $x_1$  为不可行点; 设计点  $x_2 = [0.554, 1.307]^T$  对应  $f(x_2) = -3.80$ , 这个设计点满足约束条件  $g = x_1^2 - x_2 + 1 \leq 0$ , 设计点  $x_2$  为可行点。

在优化设计问题中,有一些特殊的设计解或设计点位于优化问题的约束边界上,称为边界设计点。一般情况下,优化问题的最优化解位于约束的边界上。在优化设计问题中,有一些边界设计解满足某一个约束条件,这个约束条件就可称为起作用约束条件;如果设计解不满足某一个约束条件,这个约束条件就可称为不起作用约束条件。起作用约束条件组成的集合可称起作用约束条件集。不起作用约束条件组成的集合可称不起作用约束条件集。例如点  $x^*$  在约束条件  $g(x^*)$  上,即满足  $g(x^*) = 0$ , 故  $g(x^*)$  称为起作用约束,而其他约束条件在这点上均不起作用,又称不起作用约束。

一般还可将优化设计问题按是否存在约束条件进行分类。如果优化设计问题中不含有约束条件,那么该

优化设计问题可称为无约束优化设计问题。如果优化设计问题中含有约束条件,那么该优化设计问题可称为约束优化设计问题。实际的机械优化问题中,大多是约束优化问题。针对约束优化设计问题,可依据约束条件的性质对优化设计问题进行分类。如果约束条件和目标函数都为线性约束函数,那么该约束优化设计问题可称为线性约束优化设计问题,有时可简称为线性优化问题。如果约束条件或目标函数为非线性约束函数,那么该约束优化设计问题可称为非线性约束优化设计问题,有时可简称为非线性优化问题。

### 1.3 优化算法分类

解决优化设计问题的优化算法种类十分繁多。针对线性约束优化问题和无约束优化问题(即无约束极值问题)的优化算法在理论上和方法实现上已经非常成熟。线性约束优化和无约束优化算法也是求解非线性优化问题的基础。由于非线性约束优化设计问题的目标函数或约束条件中存在设计变量的非线性函数,尽管求解的优化算法已有许多,但是仍需要在计算效率等方面继续完善。

尽管求解优化设计问题的算法很多,但仍可依据求解问题有无约束条件将优化算法分为无约束优化算法和约束优化算法二类。

无约束优化算法主要包括坐标轮换法、最速下降法、牛顿法、共轭梯度法、Powell法、变尺度法、单纯形法等。约束优化算法主要包括 Monte Carlo法、随机方向搜索法、复合形法、可行方向法、广义简约梯度法、罚函数法、序列线性规划、序列二次规划法等。

在无约束优化算法中,各种优化方法各有优缺点。坐标轮换法具有不需要导数信息的优点,计算过程也比较简单,程序实现也比较容易,但存在算法收敛速度较慢、计算效率低等缺点。坐标轮换法主要用来解决优化问题设计变量数目小于10的小规模无约束优化问题;另外,坐标轮换法还可解决目标函数的等值线为圆或平行于坐标轴的优化问题。与其他无约束优化算法相比,最速下降法具有方法简单等优点,计算效率在最初几步迭代时较高,且对初始点不敏感,因而常与其他方法一起使用,但最速下降法需要目标函数的一阶导数信息。求解无约束优化问题的牛顿法对给定的初始点比较敏。如果初始点选择的比较好,则其解决优化问题的收敛过程会很快;如果选择不当,则可能会出现收敛失败的情况。另外,牛顿法存在计算过程复杂、计算量特别大等缺点,因此主要适合于设计变量数目较小的优化问题及目标函数阶次较低的优化问题。共轭梯度法具有收敛速度快等优点,其收敛速度远快于最速下降法。共轭梯度法计算简单,所需要的存储空间少,

适合于优化变量数目较多的中等规模优化问题。在无约束优化方法中, Powell法是计算效率比较高的优化算法之一,它不需要目标函数的导数,是求解中小型规模优化问题的有效方法。变尺度法也是计算效率比较高的优化算法之一,可用来解决高阶目标函数的优化问题,但存在程序实现比较复杂、存储空间比较大等缺点。单纯形法具有不需目标函数导数信息、程序实现简单、计算效率比较高等优点。

求解约束优化问题的约束优化算法一般以非常成熟的无约束优化算法、线性规划和二次规划类优化算法为基础发展起来的。一般可将无约束优化算法分为直接法和间接法二类。所谓直接法就是在优化过程中直接考虑约束条件的优化方法,随机试验法、随机搜索法、复合形法都属于直接类优化算法;所谓间接法就是在优化过程中将约束优化问题等效转化为无约束优化问题等相对简单的优化问题,在此基础上再对相对简单的优化问题进行求解。间接法包括如下三类优化方法:①以线性规划理论为基础,将原约束优化问题转化为线性规划类问题,采用线性规划类算法来求解,主要包括可行方向法、序列线性规划、简约梯度法等;②以无约束极值理论为基础,将原约束优化问题转化为无约束优化类问题,采用无约束优化算法来求解,主要方法有内点罚函数法、外点罚函数法、混合罚函数法等;③以二次规划理论为基础,将原约束优化问题转化为二次规划类问题,采用二次规划类算法来求解,主要包括序列二次规划法等。

与无约束优化方法一样,各种约束优化方法也是特点各异: Monte Carlo法具有方法简单、不需要导数信息等优点,但存在求解高维优化问题时计算量大等不足;随机方向搜索法具有优化求解过程收敛快,但存在局部寻优的不足,因而在使用时需采用选择多个不同初始点的策略;复合形法具有程序实现简单等优点,但在解决设计变量和约束条件多的优化问题时优化效率比较低;可行方向法是解决约束优化问题的有效方法之一,适合求解中等规模优化问题,但存在程序实现复杂等不足;广义简约梯度法具有算法收敛快、计算精度高等优点,但也存在程序实现复杂等不足;罚函数优化方法包括内点法、外点法、混合法等,具有方法实现简单等优点,但存在优化过程不稳定、收敛速度较慢等缺点,适宜于解决中小规模优化问题;序列线性规划法收敛较慢,只适用于非线性程度不是很强的优化问题;序列二次规划法是收敛速度较快、优化比较有效的方法之一,比较适合于中等规模优化问题;遗传算法具有通用性强、不需要导数信息、收敛较快等优点,是近十多年出现的比较有效的优化方法。

## 2 一维无约束优化方法

在实际应用中,一维无约束优化问题比较少。但是在优化设计问题求解过程中,包含了大量的一维无约束优化问题。例如下式描述的优化迭代过程:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)} \quad (44.1-3)$$

式中  $x^{(k+1)}$  和  $x^{(k)}$  —— 分别为第  $k+1$  次和第  $k$  次迭代过程的设计点;

$s^{(k)}$  —— 第  $k$  次迭代过程的搜索方向;

$\alpha^{(k)}$  —— 第  $k$  次迭代过程的最优步长。

优化迭代过程主要完成二项工作: ①确定第  $k$  次迭代过程搜索方向  $s^{(k)}$ ; ②确定第  $k$  次迭代过程  $\alpha^{(k)}$  最优步长。在第  $k$  次迭代过程设计点  $x^{(k)}$  和方向  $s^{(k)}$  确定后, 式(44.1-3)描述的优化迭代问题变成了求  $\alpha$  最优步长的一维无约束优化问题, 即

$$f(x^{(k+1)}) = \min f(x^{(k)} + \alpha^{(k)} s^{(k)}) \quad (44.1-4)$$

式(44.1-4)描述的优化问题的目标函数只是步长  $\alpha$  的一维函数, 而通过沿给定方向  $s$  求  $\alpha$  使目标函数极小的过程称为一维搜索, 而  $\alpha^{(k)}$  为最优步长。大多数优化方法中常使用一维寻优以获得  $\alpha$  或最优方向。因此一维优化搜索为最基础的优化算法, 它直接影响优化问题求解速度。

一维搜索求  $\alpha$  步长时, 需要确定搜索区间以保证在该区间内只有一个极小值, 确定搜索区间的方法可用进退法和外推法。本节主要介绍在确定了搜索区间后进行一维优化搜索时常用的切线法、黄金分割法和插值法。

### 2.1 切线法(牛顿法)

切线法又称为牛顿法, 是常用的一维优化算法。切线法的算法思想是采用切线代替弧逐渐逼近的策略来求得极值, 需要利用函数导数信息。当函数  $f(x)$  有一阶连续导数, 且  $f''(x) > 0$  时即可用切线法来求最优步长  $\alpha^*$ 。

求最优步长  $\alpha^*$  是通过使  $f(x^{(k)} + \alpha^{(k)} s^{(k)})$  最小, 即应满足条件

$$f'(x^{(k)} + \alpha^{(k)} s^{(k)}) = 0 \quad (44.1-5)$$

然后求方程的根即得  $\alpha^*$ 。

切线法主要步骤如下:

1) 令  $y = f'(x^{(k)} + \alpha s^{(k)})$ 。

2) 由进退法可确定搜索区间  $[a, b]$ , 在其区间内选初值  $\alpha^{(0)}$ , 在点  $[a^{(0)}, f'(x^{(k)} + \alpha^{(0)} s^{(k)})]$  处作切线(对  $\alpha$  求导), 设其方程为  $y' = f''(\alpha^{(0)}) + f''(\alpha^{(0)})(\alpha - \alpha^{(0)})$ , 切线斜率为  $f''(\alpha^{(0)})$ , 过点  $(\alpha^{(0)}, f'(\alpha^{(0)}))$ ,  $y'$  切线与  $\alpha$  轴相交于  $\alpha^{(1)}$  点, 令  $y' = 0$ 。

3) 求得  $\alpha^{(1)} = \alpha^{(0)} - \frac{f'(\alpha^{(0)})}{f''(\alpha^{(0)})}$ 。  $\alpha^{(1)}$  作为的第一次迭代近似值, 计算并判断  $f'(x^{(k)} + \alpha^{(1)}s^{(k)}) = 0$ , 满足时则有  $\alpha^* = \alpha^{(1)}$ ; 若不满足则过  $[\alpha^{(1)}, f'(x^{(k)} + \alpha^{(1)}s^{(k)})]$  作的切线, 设与轴交点  $\alpha^{(2)}$ :

$$\alpha^{(2)} = \alpha^{(1)} - \frac{f'(\alpha^{(1)})}{f''(\alpha^{(1)})}$$

4) 循环迭代:

$$\alpha^{(i-1)} = \alpha^{(i)} - \frac{f'(\alpha^{(i)})}{f''(\alpha^{(i)})} \quad i=0, 1, 2, 3 \dots$$

当  $i$  足够大时, 可使  $|\alpha^{(i)} - \alpha^{(i-1)}| \leq \delta$  或  $f'(x^{(k)} + \alpha^{(i)}s^{(k)}) \leq \epsilon$ 。

应当注意, 利用切线法, 必须要求函数满足:

$$f''(x^{(k)} + \alpha^{(i)}s^{(k)}) > 0 \quad (44.1-6)$$

否则在迭代过程中会出现发散现象。

## 2.2 黄金分割法 (0.618法)

黄金分割法又称 0.618 法, 也是很常用的一维优化算法, 其思想是通过不断缩短搜索区间长度的策略来确定函数的极小点。在黄金分割法中, 每次区间缩短

比例为 0.618, 即缩短比例  $\lambda = \frac{\sqrt{5}-1}{2}$ 。

黄金分割法的主要步骤如下:

1) 在  $[a, b]$  区间内取如下值:

$$\left. \begin{aligned} \alpha_0^{(2)} &= a + 0.618(b-a) \\ \alpha_0^{(1)} &= a + 0.382(b-a) \end{aligned} \right\} \quad (44.1-7)$$

2) 计算  $f(\alpha_0^{(2)})$  和  $f(\alpha_0^{(1)})$ , 并进行比较。

i. 若  $f(\alpha_0^{(2)}) > f(\alpha_0^{(1)})$ , 则极小点不会在  $[\alpha_0^{(2)}, b]$  区间内, 故丢掉这一区间, 从而取  $[a, \alpha_0^{(2)}]$  作为新的搜索区间  $[a_1, b_1]$ , 而  $\alpha_0^{(1)}$  点正好可作为新区间内的  $\alpha_1^{(2)}$  点。

由于  $a_1 = a, b_1 = \alpha_0^{(2)} = a + 0.618(b-a)$

则

$$\begin{aligned} \alpha_1^{(2)} &= a_1 + 0.618(b_1 - a_1) \\ &= a + 0.618[0.618(b-a)] \\ &= a + 0.618^2(b-a) \\ &= a + 0.382(b-a) \\ &= \alpha_0^{(1)} \end{aligned}$$

即  $\alpha_1^{(2)}$  点与  $\alpha_0^{(1)}$  点相同, 而  $\alpha_1^{(1)}$  点则为

$$\alpha_1^{(1)} = a_1 + 0.382(b_1 - a_1)$$

ii. 若  $f(\alpha_0^{(2)}) \leq f(\alpha_0^{(1)})$ , 则极小点不会在  $[a, \alpha_0^{(1)}]$  内, 故丢掉这一区间, 取  $[\alpha_0^{(1)}, b]$  作为新区间  $[a_1, b_1]$ , 而在这一新区间内的  $\alpha_1^{(1)}$  点恰好就是原来的  $\alpha_0^{(2)}$  点。而新区间内

$$\alpha_1^{(2)} = a_1 + 0.618(b_1 - a_1)$$

对新区间内的点  $\alpha_1^{(1)}$  和  $\alpha_1^{(2)}$  作与 ii 类似的计算比较, 决

定新的区间。

iii. 反复进行 ii 步骤计算, 直至获得  $\alpha^*$  达到规定精度位置。要求区间长度  $\leq \epsilon$ , 最优点取其中点。

反复进行 ii 步骤, 在 i 和 ii 两种情形下, 都由于有原区间内的一个分割点 ( $\alpha_0^{(1)}$  或者  $\alpha_0^{(2)}$ ) 可作为下一步区间的内点, 因此均可少计算一次函数值  $f(\alpha)$ , 这正是 0.618 法的优点所在。

0.618 法的效率不是最高的, 但其计算稳定, 易于使用, 故广泛采用。

## 2.3 插值法

插值法包括二次插值法和三次插值法等, 是比较有效的一维无约束优化算法。二次插值法的算法思想是对于一维优化问题  $f(\alpha) = f(x^{(k)} + \alpha s^{(k)})$ , 可用一个二次函数  $p(\alpha)$  来近似代替  $f(\alpha)$ , 而  $p(\alpha)$  的极小点很容易求得的。三次插值法的算法思想是对于  $f(\alpha) = f(x^{(k)} + \alpha s^{(k)})$ , 可用一个三次函数  $p(\alpha)$  来近似代替  $f(\alpha)$ , 而  $p(\alpha)$  的极小点很容易求得的。本节主要介绍二次插值法的主要步骤。

二次插值法的算法步骤:

1) 设  $f(\alpha)$  的搜索区间已知为  $[a_1, a_3]$ , 取三个插值点  $\alpha_1 < \alpha_2 < \alpha_3$ , 函数值分别为  $f_1 = f(\alpha_1), f_2 = f(\alpha_2), f_3 = f(\alpha_3)$ 。从而利用拉格朗日插值构造二次函数

$$\begin{aligned} p(\alpha) &= \frac{(\alpha - \alpha_2)(\alpha - \alpha_3)}{(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)} f_1 + \\ &\quad \frac{(\alpha - \alpha_1)(\alpha - \alpha_3)}{(\alpha_2 - \alpha_1)(\alpha_2 - \alpha_3)} f_2 + \\ &\quad \frac{(\alpha - \alpha_1)(\alpha - \alpha_2)}{(\alpha_3 - \alpha_1)(\alpha_3 - \alpha_2)} f_3 \quad (44.1-8) \end{aligned}$$

2) 令  $\frac{dp}{d\alpha} = 0$ , 可求得极小点

$$\alpha^{*(1)} = \frac{(\alpha_3^2 - \alpha_2^2)f_1 + (\alpha_1^2 - \alpha_3^2)f_2 + (\alpha_2^2 - \alpha_1^2)f_3}{(\alpha_3 - \alpha_2)f_1 + (\alpha_1 - \alpha_3)f_2 + (\alpha_2 - \alpha_1)f_3} \quad (44.1-9)$$

如果目标函数是二次函数, 则  $\alpha^{*(1)}$  为所求的真正精确极值点。若  $f$  为其他非二次函数, 则  $\alpha^{*(1)}$  为极小点的近似值, 需进一步迭代。为了进一步迭代, 需在已有的四个点  $\alpha_1, \alpha_2, \alpha_3, \alpha^{*(1)}$  中选择出三个点, 作为新的  $\alpha_1, \alpha_2, \alpha_3$ , 然后再进行新的二次插值。

3) 这时需比较  $\alpha^{*(1)}$  和  $\alpha_2$  的位置及  $f(\alpha^{*(1)})$  和  $f(\alpha_2)$  的值。保证新的三个点为两头大, 中间值小。

如果  $\alpha^{*(1)} > \alpha_2$ , 即在  $\alpha_2$  的右边, 此时比较:

i. 若  $f^* > f^2$ , 则取  $\alpha_3 = \alpha^*, f_3 = f, \alpha_1, \alpha_2$  不变, 即有  $\alpha_1, \alpha_2, \alpha^{*(1)}$  为新的  $\alpha_1, \alpha_2, \alpha_3$  三点。

ii. 若  $f^* < f^2$ , 则取  $\alpha_2, \alpha^{*(1)}, \alpha_3$  为新的  $\alpha_1, \alpha_2, \alpha_3$ 。

如果  $\alpha^{*(1)} < \alpha_2$ , 即在  $\alpha_2$  的左边, 此时比较:

- i. 若  $f^* > f^2$ , 则取  $\alpha^{*(1)}, \alpha_2, \alpha_3$  为新的  $\alpha_1, \alpha_2, \alpha_3$  点。
- ii. 若  $f^* < f^2$ , 则取  $\alpha_1, \alpha^{*(2)}, \alpha_3$  为新的三点  $\alpha_1, \alpha_2, \alpha_3$ 。
- 4) 用新的  $\alpha_1, \alpha_2, \alpha_3$  代入 2), 循环迭代。

### 3 多维无约束优化算法

多维无约束优化问题是指包含多个设计变量、对设计变量取值没有直接限制的优化问题。在实际应用中多维无约束问题比较少, 但是实际应用中的多维有约束优化问题都可等效转化为无约束优化问题来求解。因此多维无约束优化算法是约束优化设计算法的基础, 是优化设计的重要组成部分。本节简单介绍适用于多维无约束优化问题的坐标轮换法、最速下降法、变尺度法等。其他多维无约束优化算法请参考有关文献。

#### 3.1 坐标轮换法

坐标轮换法是不需要函数导数的多维无约束优化方法。其基本思想是: 每次对  $n$  个设计变量中的一个变量沿其坐标轴方向进行一维搜索, 其余各变量均固定不动; 依次轮换各个坐标轴方向进行一维搜索, 直到全部坐标轴搜索完成一次为止; 然后进行第二轮坐标轮换搜索; 经过多轮搜索, 直至求得函数最小点为止。坐标轮换法的实质就是将多维优化问题轮化为一系列一维优化问题, 再利用一维优化算法来求解。

坐标轮换法主要内容就是对设计点沿着坐标轴的方向进行循环搜索。循环搜索的主要过程如下:

对于第  $k$  轮搜索, 迭代式为

$$x_i^{(k)} = x_{i-1}^{(k)} - \alpha_i^{(k)} s_i^{(k)} \quad (i=1, 2, \dots, n) \quad (44.1-10)$$

式中  $s_i^{(k)}$  ——第  $k$  轮的第  $i$  个坐标轴方向的搜索方向, 即  $s_i^{(k)} = [0 \ 0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$ ;

$x_{i-1}^{(k)}$  ——第  $i$  个方向的前一个坐标轴方向的搜索结果向量;

$\alpha_i^{(k)}$  ——步长, 可为正(坐标轴正向), 可为负(坐标轴负向)。

在坐标轮换法中, 步长可以通过随机策略、加速步长策略来确定, 也可采用一维优化算法来确定。

#### 3.2 最速下降法 (Cauchy 法, 一阶梯度法)

最速下降法又称为 Cauchy 法或一阶梯度法。最速下降法是以数学分析中极值理论(沿函数负梯度方向目标函数值下降最快)为基础, 借助目标函数的梯度信息进行设计点的迭代过程。因此最速下降法主要思

想是, 在  $x^{(k)}$  点搜索方向取该点的负梯度方向为搜索方向, 来求目标函数的极小值。 $x^{(k)}$  点的负梯度方向为  $s^{(k)} = -\frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|}$ , 这里  $s^{(k)}$  进行了归一化处理, 即为单位矢量, 因此最速下降法在  $x^{(k)}$  点的迭代公式为

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)} = x^{(k)} - \alpha^{(k)} \frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|} \quad (44.1-11)$$

最速下降法算法的步骤如下:

1) 取初始点  $x^{(0)} \in R^n$ , 设置误差精度  $\epsilon_1 > 0$ , 令  $k=0$ 。

2) 算  $\nabla f(x^{(k)})$ ,  $s^{(k)} = -\frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|}$ 。

3) 用下式判别搜索过程是否收敛:

$$\|\nabla f(x^{(k)})\| \leq \epsilon_1 \quad (44.1-12)$$

收敛判断准则也可采用式(44.1-13)或式(44.1-14)。

$$\|x^{(k)} - x^{(k-1)}\| \leq \epsilon_2 \quad (44.1-13)$$

$$\frac{|f(x^{(k)}) - f(x^{(k-1)})|}{\|f(x^{(k)})\|} \leq \epsilon_3 \quad (44.1-14)$$

4) 对式(44.1-15)目标函数进行一维搜索, 求出最优步长  $\alpha^{(k)}$

$$\min_{\alpha} f(x^{(k)} + \alpha s^{(k)}) = f(x^{(k)} + \alpha^{(k)} s^{(k)}), \alpha > 0 \quad (44.1-15)$$

5) 令  $x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}$ ,  $k=k+1$ , 转步骤 2) 进行下一轮迭代。

与其他无约束优化算法相比, 最速下降法具有如下特点:

1) 速下降法在最初几步迭代过程中函数下降较快。

2) 最速下降法在接近最优时, 迭代收敛速度较慢(搜索步长较小)。

3) 由于最速下降法中  $\nabla f(x^{(k)})$  与  $s^{(k)}$  正交, 即当前点的目标梯度与前一点的搜索方向  $s^{(k)}$  正交, 因此搜索优化解的迭代路线为锯齿形状。

#### 3.3 牛顿法

牛顿法基本思想是利用二次曲线来逐点近似原目标函数, 以二次曲线的极小点来近似原目标函数的极小点并逐渐逼近该点。取原目标函数在各迭代点附近展开的泰勒二次多项式作为每次迭代计算时用以逼近目标函数的二次曲线的函数表达式。

牛顿法对二次函数的优化问题很有效。在一般情况下, 优化函数  $f(x)$  不一定是二次函数, 因而通过牛顿法一次迭代得到的设计点不可能是  $f(x)$  的极值点。



但是由于在  $x^{(k)}$  点附近,可用二次函数来逼近  $f(x)$  所以可以用牛顿法一次迭代的设计点作为下一次迭代点  $x^{(k+1)}$ ,即构造如下迭代计算公式:

$$x^{(k+1)} = x^{(k)} - [H(x^{(k)})]^{-1} \nabla f(x^{(k)}) \quad (44.1-16)$$

牛顿法的主要步骤如下:

- 1) 选取初始点  $x^{(0)} \in R^n$ , 设置误差精度  $\epsilon, \epsilon > 0$ .
- 2) 计算  $\nabla f(x^{(k)})$  和  $[H(x^{(k)})]^{-1}$ .
- 3) 采用式 (44.1-16) 进行设计点的迭代计算.
- 4) 进行判别搜索过程是否收敛, 否则继续设计点的迭代.

### 3.4 变尺度法

变尺度法是无约束优化方法中比较重要的优化算法,是以牛顿法为基础发展而来的,所以又称拟牛顿法.变尺度法的基本思想是,通过构造一个对称正定矩阵  $A^{(k)}$  来代替牛顿法中的  $[H(x^{(k)})]^{-1}$ ,并使迭代过程中  $A^{(k)}$  逐渐逼近  $[H(x^{(k)})]^{-1}$ .这样不用求逆矩阵  $H^{-1}$  可使得迭代计算简化,而同时能保持牛顿法收敛快的优点.

变尺度法的主要迭代计算过程为

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)} = x^{(k)} - \alpha^{(k)} A^{(k)} \nabla f(x^{(k)}) \quad (44.1-17)$$

式中  $s^{(k)}$ ——搜索方向  $s^{(k)} = -A^{(k)} \nabla f(x^{(k)})$ ;

$A^{(k)}$ ——对称正定矩阵;

$\alpha^{(k)}$ ——步长,由一维搜索方法得到.

$A^{(k)}$  在迭代过程中是不断变化的,构造矩阵  $A^{(k)}$  是变尺度法的重要内容之一.

变尺度法的计算步骤如下:

- 1) 选定初始点  $x^{(0)}$ , 并给定计算精度  $\epsilon$ .
- 2) 置  $k=0, A^{(k)} = A^{(0)} = I$  (单位矩阵), 计算  $g^{(k)} = \nabla f(x^{(k)}) = \nabla f(x^{(0)})$ , 这时探索方向为  $s^{(k)} = -A^{(k)} \nabla f(x^{(k)}) = -A^{(k)} g^{(k)}$ .
- 3) 进行一维探索求  $\alpha^{(k)}$ , 使
 
$$f(x^{(k)} + \alpha^{(k)} s^{(k)}) = \min f(x^{(k)} + \alpha s^{(k)})$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}$$
- 4) 计算  $g^{(k+1)} = \nabla f(x^{(k+1)})$ , 如果  $\|g^{(k+1)}\| < \epsilon$ , 则  $x^{(k+1)}$  即为极小点, 停止迭代, 否则转下一步.
- 5) 检查迭代次数, 若  $k=n$ , 则  $x^{(0)} = x^{(n+1)}$ , 并转向步骤 2), 若  $k < n$ , 则进行下一步.
- 6) 构造新的探索方向
 
$$s^{(k+1)} = -A^{(k+1)} \nabla f(x^{(k+1)}) = -A^{(k+1)} g^{(k+1)}$$
 并令  $k=k+1$ , 转向步骤 3).

## 4 简约梯度法及广义简约梯度法

### 4.1 简约梯度法

简约梯度法 (Reduced Gradient Method) 又称 RG 法, 适合于求解具有线性约束函数的非线性优化问题.

下面介绍简约梯度法的基本思路.

假设优化模型为

$$\left. \begin{array}{l} \min f(x) \quad x \in R^n \\ \text{s.t.} \quad Ax = b \\ \quad \quad x \geq 0 \end{array} \right\} \quad (44.1-18)$$

式中  $A$ —— $m \times n$  系数矩阵, 约束个数  $m < n$ ;

$b$ —— $m$  维常数向量.

简约梯度法的基本思想是, 先设法减少  $m$  个设计变量, 然后再求解具有  $n-m$  个设计变量的优化问题, 即在经过简约的设计空间中进行优化问题的求解. 在简约设计空间进行优化求解时要利用函数的梯度, 故这种优化方法称为简约梯度法.

将优化模型中线性约束中设计变量和  $A$  矩阵分成如下子向量和子矩阵:

$$x = \begin{bmatrix} x^B \\ x^N \end{bmatrix}, A = [B \quad C]$$

式中  $x^B$ —— $m$  维向量 (基向量);

$x^N$ —— $(n-m)$  维向量 (非基向量);

$B$ —— $m \times m$  矩阵, 对应于  $x^B, x^B = [x_1, x_2, \dots, x_m]^T = [x_1^B, x_2^B, \dots, x_m^B]^T$ ;

$C$ —— $m \times (n-m)$  矩阵, 对应于  $x^N$ .

则线性约束方程可写为

$$\begin{bmatrix} B & C \end{bmatrix} \begin{bmatrix} x^B \\ x^N \end{bmatrix} = b$$

$$Bx^B + Cx^N = b \quad (44.1-19)$$

如果  $|B| \neq 0$ , 即  $B$  为满秩 (亦即非奇异) 时有

$$x^B = B^{-1}(b - Cx^N) \quad (44.1-20)$$

如果  $x^B \geq 0$  (当  $x^B < 0$  时, 由于不满足原约束函数中  $x \geq 0$  的条件, 不考虑) 则  $x^B$  才为可行解, 并且还要求  $x^N \geq 0$ , 若  $x^B > 0$  则称之为非退化基向量, 否则称退化基向量.

假定为非退化情况 (含有退化情况时, 需将其处理成非退化情况才进行求解); 则

将  $x^B$  用  $x^N$  表示 (实质上是用  $m$  个方程消去了  $x^B$  (即  $m$  个设计变量) 向量), 则目标函数可写为

$$f(x^B, x^N) = f(x^B(x^N), x^N) = F(x^N) \quad (44.1-21)$$

对这一简约空间中的  $x^N$  (非基向量) 进行求解, 采用最速下降法 (梯度法) 进行求解则称为简约梯度法.

令  $r(x^N) = -\nabla F(x^N)$ ,  $r(x^N)$  称为  $f(x)$  的简约梯度。

为求  $r(x^N)$  可先考虑  $f(x)$  的梯度

$$\nabla f(x) = [\nabla_{x^B} f(x), \nabla_{x^N} f(x)]^T \quad (44.1-22)$$

注意  $\nabla F(x^N) = r(x^N) \neq \nabla_{x^N} f(x)$ 。

因为变量  $x^B$  为  $x^N$  的函数, 根据复合函数求导, 得到

$$r(x^N) = \frac{dF}{dx^N} = \frac{\partial f}{\partial x^N} + \left( \frac{\partial x^B}{\partial x^N} \right)^T \cdot \frac{\partial f}{\partial x^B}$$

由  $x^B = B^{-1}b - B^{-1}Cx^N$  得到

$$\frac{\partial x^B}{\partial x^N} = -[B^{-1} \cdot C]^T$$

$$\therefore r(x^N) = \frac{\partial f(x^B(x^N), x^N)}{\partial x^N}$$

$$(B^{-1}C)^T \frac{\partial f(x^B(x^N), x^N)}{\partial x^B}$$

设在当前迭代点, 欲求  $x^{(k+1)}$ , 由  $x^{(k+1)} = x^{(k)} + \alpha s^{(k)}$  得到

$$x^{(k+1)} = [x^{B(k+1)} \quad x^{N(k+1)}]^T$$

$$x^{(k)} = [x^{B(k)} \quad x^{N(k)}]^T$$

取  $s^{(k)} = -r(x^{N(k)})$  (负梯度方向) 来求  $x^N$ , 则

$$x^{N(k+1)} = x^{N(k)} - \alpha r(x^{N(k)}), \alpha > 0$$

由于约束条件要求  $x^{(k+1)} \geq 0$ , 如果第  $j$  个分量为 0, 则由于  $\alpha > 0$ ,

当  $r_j(x^{N(k)}) > 0$  时, 必有  $x_j^{N(k+1)} < 0$ , 这样的点为不可行解, 因此, 不能完全采用负梯度方向来求全部的  $x^{N(k+1)}$  的分量, 为此定义简约梯度法的搜索方向

$$p^{N(k)} = \begin{cases} 0, x_j^{N(k)} = 0, \text{且 } r_j(x^{N(k)}) > 0 \\ -r_j(x^{N(k)}), \text{除上述情况外} \end{cases} \quad (j=1, 2, \dots, n-m)$$

故迭代公式为

$$x^{N(k+1)} = x^{N(k)} + \alpha p^{N(k)}$$

上述迭代公式虽然能使  $x^{N(k+1)} \geq 0$  得到满足, 但还应保证  $x^{B(k+1)} \geq 0$ 。

由  $x^B$  的计算式有

$$x^{B(k)} = B^{-1}b - B^{-1}Cx^{N(k)}$$

$$x^{B(k+1)} = B^{-1}b - B^{-1}Cx^{N(k+1)}$$

代入  $x^{N(k+1)}$  得:

$$\begin{aligned} x^{B(k+1)} &= B^{-1}b - B^{-1}C(x^{N(k)} + \alpha p^{N(k)}) \\ &= x^{B(k)} - \alpha B^{-1}Cp^{N(k)} \end{aligned}$$

令

$$p^{B(k)} = -B^{-1}Cp^{N(k)} \quad (44.1-23)$$

则有

$$x^{B(k+1)} = x^{B(k)} + \alpha p^{B(k)} \quad (44.1-24)$$

当  $x^{N(k+1)}$  和  $x^{B(k+1)}$  合并起来写成迭代格式:

$$x^{(k+1)} = x^{(k)} + \alpha p^{(k)} \quad (44.1-25)$$

$$p^{(k)} = [p^{B(k)} \quad p^{N(k)}] \quad (44.1-26)$$

迭代应保证  $x^{(k+1)} \geq 0$ , 即各分量  $x_j^{(k+1)} = x_j^{(k)} + \alpha p_j^{(k)} \geq 0, j=1, 2, \dots, n$ 。

如果  $p_j^{(k)} \geq 0$ , 则  $x_j^{(k+1)} \geq 0$  自然满足 (注意上一步迭代设定的  $x_j^{(k)} \geq 0$ ); 如果  $p_j^{(k)} < 0$ , 则应有  $\alpha \leq \frac{x_j^{(k)}}{-p_j^{(k)}}$  才能保证  $x_j^{(k+1)} \geq 0$ 。

在所有的  $n$  个分量中求:

$$\alpha_{\max} = \min \left\{ \frac{x_j^{(k)}}{p_j^{(k)}} \mid p_j^{(k)} < 0, j=1, 2, \dots, n \right\} \quad (44.1-27)$$

则有

$$0 \leq \alpha \leq \alpha_{\max} \quad (44.1-28)$$

注意这里的  $p_j^{(k)}$  当  $j > m$  时与  $j < m$  时计算公式是不同的。

简约梯度法的计算步骤如下:

1) 选择可行初始点 (输入  $n$ ), 对分子向量

$$x^{(0)} = [x^{B(0)} \quad x^{N(0)}]^T$$

其中  $x^{B(0)} > 0$  (基向量), 给定  $\epsilon$ , 令  $k=0$ 。

2) 计算  $B, C, r(x^{N(k)})$ , 按式 (44.1-21) 和式 (44.1-23) 计算  $p^{(k)}$  搜索方向。

3) 若  $\|p^{(k)}\| < \epsilon$ , 停止迭代,  $x^* = x^{(k)}$ ; 否则求  $\alpha_{\max}$ , 并在区间  $0 \leq \alpha \leq \alpha_{\max}$  内求  $\alpha^{(k)}$  使

$$f(x^{(k)} + \alpha^{(k)} p^{(k)}) = \min_{0 \leq \alpha \leq \alpha_{\max}} f(x^{(k)} + \alpha p^{(k)})$$

然后令  $x^{(k+1)} = x^{(k)} + \alpha p^{(k)}$

4) 如果  $x^{B(k+1)} \geq 0$ , 则基向量不变, 并令  $k=k+1$ , 转 2), 如果  $x^{B(k+1)}$  中有某分量  $x_j^{B(k+1)} = 0$ , 则某向量的该分量从基向量中调整到非基向量中去, 而更换非基向量  $x^{N(k+1)}$  中最大的分量  $x_i^{N(k+1)}$  构成新的基向量  $x^{B(k+1)}$  和非基向量  $x^{N(k+1)}$ , 令  $k=k+1$ , 转 2)。

## 4.2 广义简约梯度法

广义简约梯度法 (Generalized Reduced Gradient Method) 又称 GRG 法。与简约梯度法相比, 广义简约梯度法能解决具有非线性约束的非线性优化问题。

广义简约梯度法的基本思路是, 对于不等式约束条件可利用起作用约束来转化为等式约束条件, 另外还可引入松弛变量使不等式变为等式约束条件。例如: 对约束  $g_i(x) \geq 0$ , 引入非负的松弛变量  $x_{n+1}$ , 则有  $g_i(x) - x_{n+1} = 0, 0 \leq x_{n+1} < \infty$ , 松弛变量只改变约束, 不改变目标函数。

考虑如下非线性约束规划问题:

$$\left. \begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } h_i(x) = 0, i = 1, 2, \dots, m \\ l_j \leq x_j \leq u_j, j = 1, 2, \dots, n \end{aligned} \right\} \quad (44.1-29)$$

可将上式问题写成如下向量形式:

$$\left. \begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } h(x) = [h_1(x), \dots, h_m(x)]^T \\ l \leq x \leq u, l, u \in R^n \end{aligned} \right\} \quad (44.1-30)$$

与简约梯度法处理策略一样,对  $x$  划分基向量和非基向量,则有  $x^{(k)} = [x^{B(k)} \ x^{N(k)}]^T$ ,  $l$  和  $u$  也分解成相应的两部分

$$l = \begin{bmatrix} l^B \\ l^N \end{bmatrix}, u = \begin{bmatrix} u^B \\ u^N \end{bmatrix}$$

并假定满足  $l^B < x^{B(k)} < u^B$ , 且设

$$\nabla_B h_i(x^{(k)}) = \left[ \frac{\partial h_i(x^{(k)})}{\partial x_1^B}, \frac{\partial h_i(x^{(k)})}{\partial x_2^B}, \dots, \frac{\partial h_i(x^{(k)})}{\partial x_m^B} \right]^T$$

( $i = 1, 2, \dots, m$ ) 线性无关, 基向量为  $m$  维, 非基向量为  $n-m$  维。

当  $\nabla_B h(x^{(k)}) = [\nabla_B h_1(x^{(k)}), \dots, \nabla_B h_m(x^{(k)})]$  非奇异(亦即非退化)时,

$\nabla_N h(x^{(k)}) = [\nabla_N h_1(x^{(k)}), \dots, \nabla_N h_m(x^{(k)})]$  在  $x^{(k)}$  点应有

$$h(x^{B(k)}, x^{N(k)}) = 0 \text{ (约束条件)} \quad (44.1-31)$$

为求  $x^{(k+1)} = x^{(k)} + \alpha s^{(k)}$ , 可先由  $x^{(k)}$  求出  $x^{N(k+1)}$

$$x^{N(k+1)} = x^{N(k)} + \alpha p^{N(k)} \quad (44.1-32)$$

在上式中需要先确定  $p^{N(k)}$ , 为此由式(44.1-31)对  $x^N$  求导:

$$\frac{\partial h}{\partial x^N} + \frac{\partial h}{\partial x^B} \cdot \frac{\partial x^B}{\partial x^N} = 0$$

$$\text{亦即 } \frac{\partial h_i}{\partial x_j^N} + \left[ \frac{\partial x_1^B}{\partial x_j^N}, \frac{\partial x_2^B}{\partial x_j^N}, \dots, \frac{\partial x_m^B}{\partial x_j^N} \right] \begin{bmatrix} \frac{\partial h_i(x)}{\partial x_1^B} \\ \vdots \\ \frac{\partial h_i(x)}{\partial x_m^B} \end{bmatrix} = 0 \quad (44.1-33)$$

式中  $i = 1, 2, \dots, m; j = 1, 2, \dots, n-m$ 。

记  $g_i(x^N) = x_i^B, i = 1, 2, \dots, m$ , 它可由约束等式(44.1-31)直接求出:  $x^B = g(x^N)$ 。

$$\begin{aligned} \nabla g(x^N) &= \begin{bmatrix} \frac{\partial g_1(x^N)}{\partial x_1^N} & \dots & \frac{\partial g_m(x^N)}{\partial x_1^N} \\ \vdots & & \vdots \\ \frac{\partial g_1(x^N)}{\partial x_{n-m}^N} & \dots & \frac{\partial g_m(x^N)}{\partial x_{n-m}^N} \end{bmatrix} \\ &= [\nabla g_1(x^N) \dots \nabla g_m(x^N)] \end{aligned}$$

则式(44.1-33)为

$$\nabla_N h(x) + \nabla g(x^N) \nabla_B h(x) = 0 \quad (44.1-34)$$

$$\nabla g(x^N) = -\nabla_N h(x) [\nabla_B h(x)]^{-1} \quad (44.1-35)$$

计算目标函数  $f$  关于  $x^N$  (非基向量)的简约梯度(复合函数求导)

$$\begin{aligned} r_j(x^N) &= \frac{\partial f(x)}{\partial x_j^N} + \left[ \frac{\partial g_1(x)}{\partial x_j^N}, \dots, \frac{\partial g_m(x)}{\partial x_j^N} \right] \\ &\quad \times \begin{bmatrix} \frac{\partial f(x)}{\partial x_1^B} \\ \vdots \\ \frac{\partial f(x)}{\partial x_m^B} \end{bmatrix} \quad j = 1, 2, \dots, n-m \end{aligned}$$

即

$$\begin{aligned} r(x^N) &= \nabla_N f(x) + \frac{\partial f(x)}{\partial x^B} \cdot \frac{dx^B}{dx^N} \\ &= \nabla_N f(x) + \nabla g(x^N) \cdot \nabla_B f(x) \end{aligned}$$

代入  $\nabla g(x^N)$  得  $x^N$  的简约梯度为

$$r(x^N) = \nabla_N f(x) - \nabla_N h(x) [\nabla_B h(x)]^{-1} \nabla_B f(x) \quad (44.1-36)$$

将  $x = x^{(k)}$  点代入式(44.1-36)可得  $r(x^{N(k)})$ , 得到简约梯度向量  $r(x^{N(k)})$  后, 与简约梯度法中类似, 定义搜索方向  $p^{N(k)}$  为

$$p_j^{N(k)} = \begin{cases} 0, & \text{当 } x_j^{N(k)} = l_j^N, \text{ 且 } r_j(x^{N(k)}) > 0 \\ 0, & \text{或当 } x_j^{N(k)} = u_j^N \text{ 且 } r_j(x^{N(k)}) < 0 \\ -r_j(x^{N(k)}), & \text{除上述情况外} \end{cases} \quad (j = 1, 2, \dots, n-m) \quad (44.1-37)$$

故由  $p^{N(k)}$  方向可得:

$$x^{N(k+1)} = x^{N(k)} + \alpha p^{N(k)} \quad (44.1-38)$$

下面的问题是求得  $p^{N(k)}$  后, 如果采用一维搜索求  $\alpha$ , 则  $\alpha$  要能满足约束条件式(44.1-31)是非常困难的(因为约束为非线形的), 因此常采用牛顿迭代法来求  $x^B$ , 而不是用一维搜索法求  $\alpha$  后再求  $x^B$ 。具体计算步骤如下(求  $x^B$ ):

取某一合适的步长  $\alpha'$ , 使

$$x^{N(k+1)} = x^{N(k)} + \alpha' p^{N(k)}$$

并且  $x^{N(k+1)}$  满足  $l^N \leq x^{N(k+1)} \leq u^N$ 。当  $x^N$  决定后, 在  $x^{N(k+1)}$  和  $x^{B(k)}$  附近将  $h(x) = 0$  作线性展开, 设

$$\begin{cases} h_1(x^{B^{(k)}, x^{N^{(k+1)}}}) + (x_1^B - x_1^{B^{(k)}}) \frac{\partial h_1(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x_1^B} \dots + (x_m^B - x_m^{B^{(k)}}) \frac{\partial h_1(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x_m^B} = 0 \\ \dots \\ h_m(x^{B^{(k)}, x^{N^{(k+1)}}}) + (x_1^B - x_1^{B^{(k)}}) \frac{\partial h_m(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x_1^B} \dots + (x_m^B - x_m^{B^{(k)}}) \frac{\partial h_m(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x_m^B} = 0 \end{cases}$$

上述  $m$  阶线性方程组写成矩阵形式为

$$h(x^{B^{(k)}, x^{N^{(k+1)}}}) + \frac{\partial h(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x^B} \cdot \Delta x^B = 0 \quad (44.1-39)$$

式中,  $\Delta x^B = [(x_1^B - x_1^{B^{(k)}}), \dots, (x_m^B - x_m^{B^{(k)}})]^T$

由式(44.1-39)解得:

$$\Delta x^B = - \left[ \frac{\partial h(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x^B} \right]^{-1} \cdot h(x^{B^{(k)}, x^{N^{(k+1)}}})$$

$$\text{即 } x^{B^{(k+1)}} = x^{B^{(k)}} - \left[ \frac{\partial h(x^{B^{(k)}, x^{N^{(k+1)}}})}{\partial x^B} \right]^{-1} \cdot h(x^{B^{(k)}, x^{N^{(k+1)}}}) \quad (44.1-40)$$

通过反复迭代求得  $x^{B^{(k+1)}}$ , 使之满足  $h(x^{B^{(k+1)}, x^{N^{(k+1)}}}) = 0$ , 得到  $x^{B^{(k+1)}}$  后, 判断是否满足

$$\begin{aligned} f(x^{B^{(k+1)}, x^{N^{(k+1)}}}) &< f(x^{B^{(k)}, x^{N^{(k+1)}}}) \\ l^B \leq x^{B^{(k+1)}} \leq u^B \end{aligned}$$

满足则得新点  $x^{(k+1)} = \begin{bmatrix} x^{B^{(k+1)}} \\ x^{N^{(k+1)}} \end{bmatrix}$ ; 若不满足, 则应缩小

步长重新求新的  $x^{N^{(k+1)}}$  和新的  $x^{B^{(k+1)}}$  当步长  $\alpha$  适当小时, 只要  $h(x) = 0$  有解, 则总可找到一个新的  $x^{(k+1)}$ 。

下面介绍广义简约梯度的算法主要步骤。广义简约梯度法主要步骤如下:

1) 选初始可行点  $x^{(0)}$ , 划分基向量和非基向量  $x^{(0)}$

$$= \begin{bmatrix} x^{B^{(0)}} \\ x^{N^{(0)}} \end{bmatrix}, \text{给定误差 } \varepsilon_1, \varepsilon_2, \text{令 } k=0.$$

2) 求非基向量的简约梯度  $r(x^{N^{(k)}})$  及搜索方向  $p^{N^{(k)}}$ , 如果  $\|p^{N^{(k)}}\| < \varepsilon_1$ , 停止迭代,  $x^* = x^{(k)}$ , 否则转 3)。

3) 取步长  $\alpha > 0$ , 计算  $x^{N^{(k+1)}}$ ;  $x^{N^{(k+1)}} = x^{N^{(k)}} + \alpha p^{N^{(k)}}$ , 若  $l^N \leq x^{N^{(k+1)}} \leq u^N$  则转 4); 否则缩小步长, 令  $\alpha = \frac{1}{2}\alpha$ , 从新计算  $x^{N^{(k+1)}}$  直至  $x$  的上下界约束设定满足后再转 4)。

4) 牛顿迭代求解非线性方程组的  $x^{B^{(k+1)}}$ ;  $x^{(k+1)} = [x^{B^{(k+1)}, x^{N^{(k+1)}}}]^T$ , 牛顿迭代收敛精度为  $\|h(x^{(k+1)})\| < \varepsilon_2$ , 然后判断是否满足

$$\begin{aligned} f(x^{B^{(k+1)}, x^{N^{(k+1)}}}) &< f(x^{(k)}) \\ l^B \leq x^{B^{(k+1)}} \leq u^B \end{aligned}$$

满足时转 5), 否则缩小步长  $\alpha \left( \alpha = \frac{1}{2}\alpha \right)$  转 3)。  $x^{(k+1)} = [x^{B^{(k+1)}, x^{N^{(k+1)}}}]^T$ , 令  $k=k+1$ , 判断基向量是否改变, 必要时换基(即当某个分量等于上界或下界时, 将该变量换出, 而非基向量中选一分量进入基向量), 得新的基向量  $x^B, x^N$  (非基向量), 然后转入 2)。广义简约梯度法实质上是在  $x^{(k)}$  点处, 沿起作用约束集的线性近似边界上进行搜索。

5) 在从  $x^{(k)}$  点出发搜索时, 得到的中间过程点  $[x^{N^{(k+1)}, x^B}]^T$  离开了可行区域时, 应调整  $x^B$  点使之位于约束边界  $h(x) = 0$  上, 即得  $[x^{N^{(k+1)}, x^{B^{(k+1)}}}]^T$ 。当求不出满足条件的  $x^{B^{(k+1)}}$  时需改变  $x^{N^{(k+1)}}$  再调整  $x^B$ 。

## 5 罚函数法

罚函数法 (Sequential Unconstrained Minimization Technology, SUMT) 是一种序列无约束优化方法, 因此可称为序列无约束优化技术, 有时简称为 SUMT。罚函数法虽然是序列无约束优化方法, 但与其他序列无约束优化方法 (例如简约梯度法和广义简约梯度法) 不同。简约梯度法和广义简约梯度法的基本思想是沿可行域的边界 (或线性近似边界) 搜索约束最优点, 但是当约束条件的非线性程度极高时, 这种沿边界搜索是困难的。罚函数法基本思路是采用将约束优化问题转化为无约束优化问题的策略, 利用现有成熟的多维无约束优化方法来解决。

在罚函数法中, 需要根据约束函数来构造新的目标函数, 将约束非线性规划优化问题转化为一系列无约束非线性规划优化问题。对于违反约束的迭代点通过罚因子的形式来赋予很大的目标函数值, 从而迫使搜索点逐步满足所有的约束条件。

罚函数法包括内点法、外点法、混合法和增广乘子法等方法。

### 5.1 内点法

罚函数内点法适用于处理不等式约束的非线性优化问题, 不能用于等式约束的非线性优化问题。内点法中构造的新的无约束目标函数 (即罚函数) 定义在可行域内, 并在可行域内求罚函数的极值点。

对于如下优化问题:

$$\begin{aligned} \min. & f(x) \\ \text{s.t.} & g_u(x) \leq 0 \quad (u=1, 2, \dots, m) \end{aligned} \quad (44.1-41)$$

将约束条件引入到新的目标函数中,构造如下新的优化问题:

$$\min p(x, r^{(k)}) \approx \min \left\{ f(x) - r^{(k)} \sum_{u=1}^m \frac{1}{g_u(x)} \right\} \quad (44.1-42)$$

上式属无约束优化问题。上式中,  $r^{(k)}$  为第  $k$  次迭代的罚因子, 它为一递减的正数序列。递减的正数序列如下式所示:

$$r^{(0)} > r^{(1)} > r^{(2)} > \dots \rightarrow 0 \quad (44.1-43)$$

下面分析  $r^{(k)}$  对设计点的影响:

1) 如果  $x$  在可行域内, 则有  $-r^{(k)} \sum_{u=1}^m \frac{1}{g_u(x)} \geq 0$ 。

2) 如果  $x$  在可行域内,  $p(x, r^{(k)})$  逐渐靠近边界  $g_u(x) = 0$  时, 则上述惩罚项急剧增大直至无穷大, 此时罚函数趋于无穷大。

3)  $p(x, r^{(k)})$  的第二项使约束成为搜索过程中的一个障碍以限制搜索跳出可行域范围, 故该项又称障碍函数或阻碍项。

4) 当  $\lim_{k \rightarrow \infty} r^{(k)} = 0$  时, 则  $p(x, r^{(k)})$  的极值点与约束优化问题的极值点 ( $f(x)$  的极值点) 相同。

5) 如果把  $r^{(k)}$  看作参数, 对应每一次 ( $k$ ), 令搜索到的无约束极值的最优解为  $x^*(r^{(k)})$ , 将  $x^*(r^{(k)})$  看作以  $r^{(k)}$  为参数的轨迹, 当参数  $r^{(k)} \rightarrow 0$  (即  $k \rightarrow \infty$ ) 时,  $x^*(r^{(k)})$  沿该参数轨迹收敛于  $f(x)$  的最优点, 而这种逼近收敛是从可行域内部来逼近的。事实上是逼近选取  $r^{(k+1)} = cr^{(k)}$  来获取不同的  $x^*$ 。

6) 对应每个  $r^{(k)}$  (选定的参数), 以上一次优化搜索的最优点为本次优化搜索的起始点进行探索, 这时需确定搜索方向和搜索步长。

罚函数的内点法主要计算迭代步骤如下:

1) 取初始罚因子  $r^{(0)} > 0$ , 允许误差  $\epsilon > 0$ , 输入  $n$ ,  $c$  (递减系数)。

2) 在可行域内选可行初始点  $x^{(0)}$ , 令  $k=0$ 。

3) 求无约束优化问题  $\min_x p(x, r^{(k)})$ ,  $x \in D$  (可行域) 的极值点  $x^*(r^{(k)})$ 。初始点为  $x^{(0)}$ ,  $r^{(k)}$  已知, 定出搜索方向和搜索步长即可。

4) 进行收敛判断。如果  $\|x^*(r^{(k)}) - x^*(r^{(k-1)})\| \leq \epsilon_1 = 10^{-3} \sim 10^{-7}$ , 且  $\left\| \frac{p(x^*, r^{(k)}) - p(x^*, r^{(k-1)})}{p(x^*, r^{(k-1)})} \right\| \leq \epsilon_2 = 10^{-3} \sim 10^{-4}$  停止迭代, 取  $x^* = x^*(r^{(k)})$ , 否则转 5)。

5)  $r^{(k+1)} = cr^{(k)}$ ,  $x^{(0)} = x^*(r^{(k)})$ ,  $k = k + 1$ , 转 3)。

式中  $c$  — 递减系数,  $c = 0.02 \sim 0.5$ , 常用  $0.1 \sim 0.5$ 。

罚函数的内点法属于经验性较强的优化方法。使用内点法时需要特别注意如下问题:

(1) 初始点  $x^{(0)}$  的选择

内点法要求初始点  $x^{(0)}$  在可行域内部并避免在边界上, 即满足  $g_u(x^{(0)}) < 0$  ( $u=1, 2, \dots, m$ )。因此可用随机法选择初始点, 或者利用罚函数对不满足约束条件的约束函数求极值法。选择初始点的方法如下:

1) 任取一点  $x^{(0)} \in R^n$ ,  $x^{(0)} > 0$ ,  $k=0$ 。

2) 求出  $g_u(x^{(k)}) < 0$  和  $g_u(x^{(k)}) \geq 0$  的指标集合  $I^{(k)}$  和  $U^{(k)}$ 。

3) 当  $U^{(k)} = \emptyset$  时停止迭代。使  $x^{(k)}$  为初始点, 否则转 4)。

4) 以  $x^{(k)}$  为初始点构造罚函数如下:

$$\min \left\{ \sum_{u \in U^{(k)}} g_u(x) - r^{(k)} \sum_{u \in I^{(k)}} \frac{1}{g_u(x)} \right\}$$

求其最优解  $x^{(k+1)}$ , 转 5)。

5) 令  $r^{(k+1)} = cr^{(k)}$ ,  $k = k + 1$ , 转 2)。

(2) 初始罚因子  $r^{(0)}$  的选择

在内点法中,  $r^{(0)}$  的选择直接影响计算效率, 如果  $r^{(0)}$  太小, 容易使罚项不足够大, 迭代时可能会超出可行域, 如果  $r^{(0)}$  太大, 则最初无约束极值点离边界太远, 计算效率太低, 一般可使罚项 (实质是阻碍函数) 与  $f(x)$  相等, 即

$$r^{(0)} = \frac{f(x^{(0)})}{\sum_{u=1}^m \frac{1}{g_u(x^{(0)})}} \quad (44.1-44)$$

但上式对于在  $x^{(0)}$  接近边界情况下, 计算出的  $x^{(0)}$  太小, 故可以增加一个新的约束条件:

$$g_{m+1}(x) = f(x) - f(x^{(0)}) \leq \epsilon \quad (44.1-45)$$

它保证所选  $r^{(0)}$  不会使  $f(x)$  比  $f(x^{(0)})$  大太多。再构造如下罚函数:

$$p(x, r^{(k)}) = f(x) - r^{(k)} \times \left\{ \sum_{u=1}^m \frac{1}{g_u(x)} + \frac{1}{f(x) - f(x^{(0)}) - \epsilon} \right\} \quad (44.1-46)$$

上式中罚函数可使  $r^{(0)}$  较大时, 第一步搜索的无约束极值点即使离可行域边界太远, 对计算效率的影响也不大。

罚函数的内点法优点在于求解过程能产生逐步改善的可行解, 这点与可行方向法是类似的。

## 5.2 外点法

与内点法将罚函数定义在可行域内不同, 外点法是将罚函数定义在可行域外。因此求解罚函数的无约束问题时, 将从可行域外逐步逼近无约束优化问题的优化解。

构造的外点罚函数为:

$$p(x, r^{(k)}) = f(x) + r^{(k)} \sum_{u=1}^m [\max(0, g_u(x))]^2 \quad (44.1-47)$$

式中 右边第二项 —— 罚项;

$\alpha$  —— 罚函数指数, 一般  $\alpha=2$ ;

$r^{(k)}$  —— 罚因子, 为大于零的递增序列  
递增序列如下:

$$0 < r^{(0)} < r^{(1)} < \dots < r^{(k)}$$

且  $\lim_{k \rightarrow \infty} r^{(k)} \rightarrow +\infty$

罚函数中的障碍函数为

$$\max(0, g_u(x)) = \begin{cases} g_u(x) & g_u(x) > 0 \\ 0 & g_u(x) \leq 0 \end{cases} \quad (44.1-48)$$

通过障碍函数确保迭代点在可行域外。外点法罚函数与下式优化问题等价,

$$p(x, r^{(k)}) = \begin{cases} f(x) + r^{(k)} \sum_{u=1}^m [g_u(x)]^2 & \text{可行域外} \\ f(x) & \text{可行域内} \end{cases} \quad (44.1-49)$$

如果原问题具有等式约束, 则式中也包括等式约束(具体体现在衰减函数中)。

$$p(x, r^{(k)}) = \begin{cases} f(x) + r^{(k)} \sum_{u=1}^m [g_u(x)]^2 & \text{可行域外} \\ + r^{(k)} \sum_{v=1}^p [h_v(x)]^2 & \\ f(x) & \text{可行域内} \end{cases} \quad (44.1-50)$$

外点法是从可行域外部向原目标函数的约束最优点逼近直至达到最优点。在  $p(x, r^{(k)})$  中带罚因子的项(罚项)的值  $r^{(k)} \sum_{u=1}^m [g_u(x)]^2$  在迭代过程中逐渐减小, 故称其为衰减函数, 注意尽管  $r^{(k)}$  逐渐增大, 但是由于在逐渐接近最优点时,  $g_u(x)$  接近于 0, 故衰减函数是减小的。

当  $g_u(x)$  违反约束越多时,  $r^{(k)}$  应越大, 即惩罚越严重, 从而使  $p$  目标函数值越大, 这样可引导无约束优化问题  $p$  寻优时避开这一搜索范围。

外点法迭代计算步骤如下:

1) 选取  $r^{(0)} > 0$ ,  $\epsilon_1$  和  $\epsilon_2$  控制罚因子递增的系数, 令  $k=0$ , 并任意选择一初始点  $x^{(0)}$  (可行域外或内均可)。如果原始约束优化问题的极值点在可行域内部, 即约束极值点等于无约束极值点, 则罚函数优化问题始终满足  $p(x, r^{(k)}) = f(x)$  (惩罚项均为 0)。因此罚函数为无约束优化问题。如果原问题极值点在可行域边界上 (约束极值点不等于无约束极值点), 则罚函数  $p(x,$

$r^{(k)}$  的极值点在可行域外部, 一般工程问题无论  $x^{(0)}$  取的是否在可行域内, 外点罚函数法的极值点均在可行域外。

2) 从  $x^{(k)}$  点出发, 求如下无约束优化问题(设得到的极值点为  $x^*(r^{(k)})$ ):

$$\min_{x \in R^n} p(x, r^{(k)})$$

$$\min_{x \in R^n} [f(x) + r^{(k)} \sum_{u=1}^m [\max(0, g_u(x))]^2] \quad (44.1-51)$$

3) 计算  $x^*$  点违反约束的最大量  $Q = \max(g_u(x^*(r^{(k)})), u=1, 2, \dots, m)$ 。

4) 进行迭代收敛准则判别。如满足  $Q \leq \epsilon$ ,  $10^{-8} \sim 10^{-4}$ , 即已接近约束边界, 则停止迭代, 取  $x^* = x^*(r^{(k)})$ 。

5) 判断  $r^{(k)} > R$ , 若上式成立时, 停止迭代; 若  $r^{(k)} \leq R$  或二次无约束罚函数极值点距离不足够小, 则取  $r^{(k+1)} = cr^{(k)}$ ,  $x^{(0)} = x^*(r^{(k)})$ ,  $k=k+1$ , 转 2)。

应该注意到, 外点法严格说来所求的最优点仍为外点。如果优化需要严格在可行设计域内优化问题, 应将原约束边界向可行域内平移一些距离来保证最优点为内点。

外点法对初始点无可行性要求(而内点法则有), 但只在最后步骤得到最优解。内点法初始点选择困难, 不能用于等式约束, 但每个迭代点都是可行点。

### 5.3 混合法

混合法是在结合罚函数法内点法和外点法的优点基础上将内点法和外点法罚项函数结合起来构造的优化方法。混合法通过构造新的罚函数作为无约束优化问题的目标函数, 又用来求解外点法类似的问题, 采用成熟的无约束优化方法来解决约束优化问题。

定义优化问题为

$$\left. \begin{aligned} \min. & f(x) \quad x \in R^n \\ \text{s. t.} & g_u(x) \leq 0 \quad (u=1, 2, \dots, m) \\ & h_v(x) = 0 \quad (v=1, 2, \dots, p) \end{aligned} \right\} \quad (44.1-52)$$

构造如下混合罚函数:

$$p(x, r^{(k)}, m^{(k)}) = f(x) - r^{(k)} \sum_{u=1}^m \frac{1}{g_u(x)} + m^{(k)} \sum_{v=1}^p [h_v(x)]^2 \quad (44.1-53)$$

上式中, 罚函数中既有障碍函数(第二项)又有衰减函数(第三项), 一般取  $m^{(k)} = \frac{1}{\sqrt{r^{(k)}}}$ 。

这样罚因子只含一个参数, 即  $r^{(k)}$ ,  $r^{(k)}$  为递减的,

即

$$r^{(0)} > r^{(1)} > r^{(2)} > \dots \rightarrow 0$$

上式中初始点  $x^{(0)}$  应取为内点。迭代过程也同样与内点法相似。计算步骤请参考内点法有关内容。

#### 5.4 增广拉格朗日乘子法

在罚函数优化方法中,内点法和外点法在  $r^{(k)} \rightarrow 0$  (内点法) 和  $r^{(k)} \rightarrow \infty$  (外点法) 时才能真正收敛于极值点,在迭代过程中愈接近最优值时收敛愈慢,因此可以采用增广拉格朗日乘子法来克服这种收敛慢及不稳定的缺点。

增广拉格朗日乘子法通过构造如下增广拉格朗日函数(只考虑不等式约束)来克服上述缺点:

$$L(x, \lambda, r^{(k)}) = f(x) - \sum_{u=1}^m \lambda_u (g_u(x) + \omega_u^2) + r^{(k)} \sum_{u=1}^m (g_u(x) + \omega_u^2)^2 \quad (44.1-54)$$

式中  $\omega_u$ ——第  $u$  个不等式约束的松弛变量。

$$g_u(x) + \omega_u^2 = 0, u = 1, 2, \dots, m$$

由于增广拉格朗日函数中引入了松弛变量  $\omega_u$ , 使设计空间维数增加。如果  $m$  很大时, 则维数增加很大, 故可利用  $\frac{\partial L}{\partial \omega_u} = 0$ , 求  $L$  关于  $\omega$  的极小值, 通过导出  $\omega_u$  的计算式以减少维数。

通过求导, 得到

$$\begin{aligned} 2r^{(k)}(g_u(x) + \omega_u^2)2\omega_u - \lambda_u 2\omega_u &= 0 \\ 2\omega_u [2r^{(k)}(g_u(x) + \omega_u^2) - \lambda_u] &= 0 \\ 2\omega_u [2r^{(k)}\omega_u^2 - (-\lambda_u + 2r^{(k)}g_u(x))] &= 0 \quad (44.1-55) \\ \omega_u^2 &= -\frac{-\lambda_u + 2r^{(k)}g_u(x)}{2r^{(k)}} \end{aligned}$$

如果  $-\lambda_u + 2r^{(k)}g_u(x) \geq 0$ , 需满足  $\omega_u = 0$ 。

反之用上式计算  $\omega_u$ , 故设

$$g_u(x) + \omega_u^2 = \begin{cases} g_u(x) - \lambda_u + 2r^{(k)}g_u(x) \geq 0 \\ \frac{\lambda_u}{2r^{(k)}} - \lambda_u + 2r^{(k)}g_u(x) < 0 \end{cases} \quad (44.1-56)$$

上式等价于:

$$g_u(x) + \omega_u^2 = \begin{cases} g_u(x) & g_u(x) \geq \frac{\lambda_u}{2r^{(k)}} \\ \frac{\lambda_u}{2r^{(k)}} & g_u(x) < \frac{\lambda_u}{2r^{(k)}} \end{cases} \quad (44.1-57)$$

$$\text{即} \quad g_u(x) + \omega_u^2 = \max \left[ g_u(x), \frac{\lambda_u}{2r^{(k)}} \right]$$

利用此式可消去松弛变量  $\omega_u$ , 增广拉格朗日函数为

$$L(x, \lambda, r^{(k)}) = f(x) - \sum_{u=1}^m \lambda_u \varphi_u + r^{(k)} \sum_{u=1}^m \varphi_u^2 \quad (44.1-58)$$

$$\text{式中} \quad \varphi_u = \max \left[ g_u(x), \frac{\lambda_u}{2r^{(k)}} \right]$$

这样原不等式约束的优化问题转化为了上式描述的无约束优化问题, 其中拉格朗日乘子按如下迭代公式计算:

$$\lambda_u^{(k+1)} = \lambda_u^{(k)} - 2r^{(k)}\varphi_u \quad u = 1, 2, \dots, m \quad (44.1-59)$$

明显增广拉格朗日乘子法比外点法多了一项  $\sum \lambda_u g_u(x)$ , 使之在  $r^{(k)}$  为极限值时即可收敛约束极值点。迭代过程中  $r^{(k)}$  值的变化与外点法相类似, 渐步增大。

#### 6 序列线性规划法

序列线性规划法主要思路是采用一系列线性规划问题去逼近非线性规划优化问题, 然后采用成熟的线性规划优化算法来解决。序列线性规划法的线性化策略包括二种: 一种是对线性约束的非线性规划优化问题的线性化逼近; 另一种是对非线性约束的非线性规划优化问题的线性化逼近。

先介绍线性约束的非线性规划问题的线性逼近策略。

考虑如下规划优化问题:

$$\begin{aligned} \min. & f(x) \\ \text{s. t.} & Ax \leq b \\ & Ex \leq e \end{aligned} \quad (44.1-60)$$

上述问题的可行域为一凸多面体, 在可行域中的  $x^{(0)}$  点处对  $f(x)$  作线性泰勒展开,  $x^{(0)}$  称为  $f(x)$  的线性化点。在  $x^{(0)}$  点处对  $f(x)$  作线性泰勒展开

$$f(x) \approx f(x^{(0)}) - f(x^{(0)}) + [\nabla f(x^{(0)})]^T [x - x^{(0)}] \quad (44.1-61)$$

然后求解如下线性规划优化问题:

$$\begin{aligned} \min. & f'(x) \quad x \in D \quad (\text{可行域}) \quad (44.1-62) \\ \text{s. t.} & Ax \leq b \\ & Ex \leq e \end{aligned}$$

把式(44.1-61)代入上式中目标函数, 可将上式中线性规划优化问题等价于如下优化问题,

$$\min [ [\nabla f(x^{(0)})]^T x \quad x \in D \quad (44.1-63)$$

当可行域  $D$  有界时, 上式优化问题的优化解必定存在, 设其为  $y^{(0)}$ 。

1) 如果  $[\nabla f(x^{(0)})]^T [y^{(0)} - x^{(0)}] = 0$ , 则  $x^{(0)}$  为式(44.1-60)的最优解。这里  $[y^{(0)} - x^{(0)}]$  为搜索方向  $s$ , 表明不存在可行下降方向了,  $x^{(0)}$  为  $K-T$  最优值。

2) 如果上式不存在, 必有

$[\nabla f(x^{(0)})]^T [y^{(0)} - x^{(0)}] < 0, s = [y^{(0)} - x^{(0)}]$  为可行下降方向。

而根据线性规划理论可知,  $y^{(0)}$  必为  $D$  的顶点。故由

$x^{(0)}$ 出发作一维搜索步长应为  $0 \leq \alpha \leq 1$ , 即搜索范围不能超过顶点  $y^{(0)}$  与点  $x^{(0)}$  之间。

$$\min f(x^{(0)}) + \alpha(y^{(0)} - x^{(0)}) - f(x^{(0)}) + \alpha_0(y^{(0)} - x^{(0)})$$

令新点  $x^{(1)} = x^{(0)} + \alpha_0(y^{(0)} - x^{(0)})$ , 由于  $D$  为凸集, 故  $x^{(1)} \in D$ , 而  $x^{(1)}$  为  $x^{(0)}$  和  $y^{(0)}$  连线上的极小点, 求得  $x^{(1)}$  后, 将  $x$  看成  $x^{(0)}$  进行新一轮的泰勒线性展开继续求解。

下面介绍线性约束的非线性规划问题的线性逼近策略的迭代步骤。主要计算步骤如下:

- 1) 给定初始点  $x^{(0)} \in D, \epsilon$ , 令  $k=0$ 。
- 2) 计算  $\nabla f(x^{(k)})$ , 若  $\|\nabla f(x^{(k)})\| < \epsilon$ , 则  $x^* = x^{(k)}$ , 停止迭代, 否则转 3)。
- 3) 求解如下线性规划优化问题:

$$\min_{x \in D} [\nabla f(x^{(k)})]^T x$$

得最优解  $y^{(k)}$ 。

- 4) 若  $[\nabla f(x^{(k)})]^T [y^{(k)} - x^{(k)}] < \epsilon$ , 停止迭代,  $x^* = x^{(k)}$ , 否则转 5)。

- 5) 沿  $(y^{(k)} - x^{(k)})$  方向作一维搜索求:

$$\min_{0 \leq \alpha_k \leq 1} f(x^{(k)}) + \alpha(y^{(k)} - x^{(k)}) = f(x^{(k)}) + \alpha_0(y^{(k)} - x^{(k)}) \text{ 的最优步长 } \alpha_k$$

- 6) 令  $x^{(k+1)} = x^{(k)} + \alpha_0(y^{(k)} - x^{(k)})$ ,  $k = k + 1$ , 转到 2)。

对于非线性约束非线性规划问题的线性逼近法策略, 与上述问题基本类同。

假设包含非线性约束的非线性规划优化问题如下式所示:

$$\left. \begin{aligned} \min. & f(x) \\ \text{s. t.} & \left. \begin{aligned} g_u(x) &\leq 0 \quad u=1, 2, \dots, m \\ h_v(x) &= 0 \quad v=1, 2, \dots, p \\ l &\leq x \leq u \end{aligned} \right\} \end{aligned} \right. \quad (44.1-64)$$

对优化问题中的目标和约束函数在  $x^{(k)}$  点作线性泰勒展开, 得到

$$\left. \begin{aligned} \min. & f(x^{(k)}) + [\nabla f(x^{(k)})]^T [x - x^{(k)}] \\ \text{s. t.} & \left. \begin{aligned} g_u(x^{(k)}) + [\nabla g_u(x^{(k)})]^T [x - x^{(k)}] &\leq 0 \\ &u=1, 2, \dots, m \\ h_v(x^{(k)}) + [\nabla h_v(x^{(k)})]^T [x - x^{(k)}] &= 0 \\ &v=1, 2, \dots, p \\ l &\leq x \leq u \end{aligned} \right\} \end{aligned} \right. \quad (44.1-65)$$

因此求解上式中优化问题后就可求解非线性约束非线性规划优化问题。假设上式中优化问题极小点为  $x^{(k+1)}$ , 然后在  $x^{(k+1)}$  对原优化问题的目标和约束函数作线性近似, 重新求解近似后的优化问题, 一直继续上述迭代过程, 直至满足收敛条件为止。

## 7 序列二次规划法

在优化问题中如果目标函数为二次函数, 约束条件是线性的, 这类问题称为二次规划问题。目前二次规划问题已有许多成熟的优化方法, 因此二次规划优化方法可作为解决非线性优化问题的基础。序列二次规划法就是将非线性优化问题转化为一系列二次规划问题, 然后采用成熟的二次规划优化方法来求解。

采用序列二次规划法, 就是将一般的非线性规划问题 (约束函数和目标函数都为非线性函数的优化问题) 在迭代点  $x^{(k)}$  处可构造如下二次规划问题 (在  $x^{(k)}$  点作二次展开): 对目标函数作二次展开, 对约束函数作线性展开 (在  $x^{(k)}$  点)。

$$\begin{aligned} \min. & [\nabla f(x^{(k)})]^T s + \frac{1}{2} s^T H(x^{(k)}) s \quad s \in R^n \\ \text{s. t.} & h_i(x^{(k)}) + [\nabla h_i(x^{(k)})]^T s = 0 \quad i=1, 2, \dots, p \end{aligned} \quad (44.1-66)$$

$$g_u(x^{(k)}) + [\nabla g_u(x^{(k)})]^T s \leq 0 \quad u=1, 2, \dots, m$$

式中  $s$ ——待求搜索向量 (可行下降方向);

$H(x^{(k)})$ ——海色矩阵 (拉格朗日问题的)。

$$L(x, \lambda) = f(x) + \sum_{i=1}^{p+m} \lambda_i g_i(x)$$

式中  $g_i(x)$ ——等式和不等式约束集合。

$$H(x^{(k)}) = \nabla_x^2 L(x^{(k)}, \lambda)$$

$$= \nabla^2 f(x^{(k)}) + \sum_{i=1}^{p+m} \lambda_i \nabla^2 g_i(x^{(k)})$$

用  $g(x)$  表示  $g_i(x)$ , 令

$$H(x^{(k)}) = \nabla^2 f(x^{(k)}) + [\lambda^{(k)}]^T \nabla^2 g(x^{(k)}) \quad (44.1-67)$$

对二次展开的优化问题求解搜索方向 ( $s$ ) 后, 沿该方向进行一维搜索, 可得搜索步长  $\alpha$ , 从而得新点  $x^{(k+1)}$ , 然后在  $x^{(k+1)}$  根据优化问题构造新的二次规划子问题。这一系列二次规划子问题的解最后逼近原问题的解。因此这种优化方法被称为序列二次规划法。

下面介绍序列二次规划算法的主要原理

不失一般性, 考虑等式约束非线性规划问题

$$\left. \begin{aligned} \min. & f(x) \quad x \in R^n \\ \text{s. t.} & h_i(x) = 0, i=1, 2, \dots, m \end{aligned} \right\} \quad (44.1-68)$$

约束写成向量形式为  $\text{s. t. } h(x) = 0 \quad x \in R^n$

上述问题最优解存在的必要条件是拉氏函数一阶导数为 0, 故令

$$\nabla_x L(x, \lambda) = \nabla f(x) + \lambda^T \nabla h(x) = 0 \quad (44.1-69)$$

$$\nabla_x L(x, \lambda) = h(x) = 0$$

为从式 (44.1-69) 中求出极值点  $x^*$ ,  $\lambda^*$ , 需求解  $m+n$  个非线性方程组, 这可用牛顿迭代法来求解。即



在  $x^{(k)}$  点处对式 (44.1-69) 进行线性化可得:

$$\begin{cases} \nabla f(x^{(k)}) + [\lambda^{(k)}]^T \nabla h(x^{(k)}) + (x^{(k+1)} - x^{(k)})^T \\ \times [\nabla^2 f(x^{(k)}) + [\lambda^{(k)}]^T \nabla^2 h(x^{(k)}) + (\lambda^{(k+1)} - \lambda^{(k)})^T \\ \times [\nabla h(x^{(k)})] = 0 \\ h(x^{(k)}) + (x^{(k+1)} - x^{(k)})^T \nabla h(x^{(k)}) = 0 \end{cases} \quad (44.1-70)$$

令

$$\begin{aligned} H(x^{(k)}, \lambda^{(k)}) &= \nabla_x^2 L(x^{(k)}, \lambda^{(k)}) \\ &= \nabla^2 f(x^{(k)}) + [\lambda^{(k)}]^T \nabla^2 h(x^{(k)}) \\ s &= x^{(k+1)} - x^{(k)} \\ r &= \lambda^{(k+1)} - \lambda^{(k)} \end{aligned}$$

则线性方程组 (44.1-70) 可表示为

$$\begin{bmatrix} H & \nabla h^T(x^{(k)}) \\ \nabla h^T(x^{(k)}) & 0 \end{bmatrix} \begin{bmatrix} s \\ r \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix} \quad (44.1-71)$$

式 (44.1-69) 的求解是非常困难的任务。一种较好的方法是通过构造一系列二次规划子问题:

$$\begin{aligned} \min. \quad & \nabla^T f(x^{(k)})s + \frac{1}{2} s^T H s \quad s \in R^n \\ \text{s. t.} \quad & h(x^{(k)}) + \nabla^T h(x^{(k)})s = 0 \end{aligned}$$

该问题的最优解应满足拉氏函数一阶导数等于 0, 即

$$\begin{cases} \nabla_x L = \nabla f(x^{(k)}) + Hs + [\lambda^{(k)}]^T \nabla h(x^{(k)}) = 0 \\ \nabla_x L = h(x^{(k)}) + \nabla^T h(x^{(k)})s = 0 \end{cases} \quad (44.1-72)$$

令  $r = \lambda'$ , 则 (44.1-72) 式可写为

$$\begin{bmatrix} H & \nabla^T h(x^{(k)}) \\ \nabla^T h(x^{(k)}) & 0 \end{bmatrix} \begin{bmatrix} s \\ r \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix} \quad (44.1-73)$$

可见式 (44.1-73) 和式 (44.1-71) 是等价的。只要这里式 (44.1-73) 中的  $s$  和  $r$  如果定义成:  $s = x^{(k+1)} - x^{(k)}$ ,  $r = \lambda^{(k+1)} - \lambda^{(k)} = \lambda'$ , 即可。

显然拉氏函数海色矩阵  $H$  求解十分重要。海色矩阵  $H$  的计算可采用下式:

$$H = \nabla^2 f(x^{(k)}) + [\lambda^{(k)}]^T \nabla^2 g(x^{(k)}) \quad (44.1-74)$$

式中对起作用约束,  $\lambda^{(k)}$  的拉氏乘子分量等于 0, 不起作用约束才有其分量大于 0。

$H$  涉及计算二阶导数, 这一计算量太大, 故与无约束问题的变尺度法一样要用近似方法来近似  $H$ 。同样对于  $H^{-1}$ 。这可以利用一个对称正定矩阵  $A^{(k)}$  来逼近  $H^{-1}$ , 具体做法同无约束变尺度法通过逐步逼近 (例如  $A^{(k+1)} = A^{(k)} + E^{(k)}$ ) 策略类似。

## 8 遗传算法

自 20 世纪 60 年代遗传算法出现以来, 遗传算法就用来解决机械设计、工程结构设计、电力系统、自动控制等许多领域的优化设计问题。本节介绍遗传算法

的这些基本思想。

无约束优化问题一般可表示成如下形式:

$$\left. \begin{aligned} \min. \quad & f(x) \\ x &= [x_1, x_2, \dots, x_n] \\ a_i &\leq x_i \leq b_i, (i=1, 2, \dots, n) \end{aligned} \right\} \quad (44.1-75)$$

式中  $n$  ——设计变量的数目;

$b_i, a_i$  ——设计变量  $x_i$  的上、下限。

利用遗传算法解决上述无约束优化问题主要包括如下内容。

### (1) 适应度函数的建立

适应度是遗传算法中描述个体性能的主要指标。一般个体适应度值越大, 个体的性能越好; 反之, 个体适应度值越小, 个体性能亦越差。在遗传算法中, 适应度的值必须是大于等于 0 的数。由于遗传算法是依据适应度的值对个体进行优胜劣汰的, 因此, 将无约束优化问题的目标函数与个体的适应度建立映射关系, 即可在群体进化过程中实现对优化问题目标函数的寻优。

一般优化问题的目标函数将随实际情况的不同而变化。如优化问题是最小值优化问题, 优化过程中目标函数将向极小值方向变化; 如优化问题是最大值优化问题, 优化过程中目标函数向极大值方向变化。而遗传算法中适应度在群体进化过程中是向极大值逼近, 因此, 必须对优化问题的目标函数进行适当的转化, 才能与优化问题的目标函数遗传算法中的适应度建立关系。

将目标函数转换成适应度函数, 一般需遵循如下两个基本原则:

1) 适应度值必须大于等于 0;

2) 优化过程中目标函数变化方向 (如向目标函数最大值变化或向最小值变化) 应与群体进化过程中适应度函数变化方向一致。

对于最小值优化问题, 可通过下式来建立与目标函数存在映射关系的适应度函数:

$$F(x) = C - f(x) \quad (44.1-76)$$

式中  $F(x)$  ——适应度函数;

$C$  ——可调参数,  $C$  的取值应使适应度函数  $F(x)$  恒大于等于 0;

$f(x)$  ——最小值优化问题的目标函数。

为确保适应度函数  $F(x)$  不小于 0, 一般常采用下式建立适应度函数:

$$F(x) = \begin{cases} C_{\max} - f(x) & \text{当 } f(x) < C_{\max} \text{ 时} \\ 0 & \text{当 } f(x) \geq C_{\max} \text{ 时} \end{cases} \quad (44.1-77)$$

式中  $C_{\max}$  ——一个可调参数,  $C_{\max}$  的取值应使适应度函数  $F(x)$  恒大于等于零。

对于最大值优化问题,可通过下式来建立与目标函数存在映射关系的适应度函数:

$$F(x) = f(x) + C \quad (44.1-78)$$

式中  $F(x)$  —— 适应度函数;

$C$  —— 可调参数,  $C$  的取值应使适应度函数  $F(x)$  恒大于 0;

$f(x)$  —— 最大值优化问题的目标函数。

为确保适应度函数  $f(x)$  不小于 0,一般常采用下式建立适应度函数:

$$F(x) = \begin{cases} C_{\min} + f(x) & \text{当 } f(x) < 0 \text{ 时} \\ f(x) & \text{当 } f(x) \geq 0 \text{ 时} \end{cases} \quad (44.1-79)$$

式中  $C_{\min}$  —— 一个可调参数,  $C_{\min}$  取值应使适应度函数  $F(x)$  恒大于 0。

目标函数与适应度函数之间的映射关系还可由其他形式的函数关系来描述。例如,利用下式的指数函数来建立目标函数与适应度的映射关系:

$$F(x) = C^{f(x)}$$

若  $f(x)$  是最大值优化问题的目标函数,  $C$  取 1.618 或 2.0;若  $f(x)$  是最小值优化问题的目标函数,  $C$  取 0.618。显然,通过上式建立的适应度函数满足目标函数与适应度函数之间转化的两个基本规则。

### (2) 设计变量与个体间的映射

设计变量与个体间的映射可通过编码来实现。编码方法一般应遵循位串定义长度最短、模式阶次最高、模式数目最大等原则。

长度为  $L$  的二进制位串与设计变量  $x_i$  之间映射关系可由下式来表示:

$$x_i = a_i + \frac{M(b_i - a_i)}{2^L - 1} \quad (44.1-80)$$

$$M = \frac{(x_i - a_i)(2^L - 1)}{b_i - a_i} \quad (44.1-81)$$

式中  $M$  —— 由二进制位串编码对应的十进制数值;

$a_i, b_i$  —— 设计变量的上下限。

从式 (44.1-80) 和式 (44.1-81) 可以看出,对应于 1 位个体的设计变量  $x_i$  实际上变成了一个离散变量,设计变量的离散间隔为  $(b_i - a_i)/(2^L - 1)$ 。当  $L$  取值越来越大,离散间隔就越来越小;当  $L$  趋于无穷大时,离散间隔就趋于 0。这时,设计变量与个体间的离散映射关系就变成了连续映射关系。

### (3) 群体初始化

群体的初始化一般包括如下内容:

1) 确定群体规模数目  $m$ , 群体规模数一般在 50 以上;

2) 对优化问题的初始解  $x_0 = [x_1, x_2, \dots, x_n]$  进行编码,产生与初始解对应的个体;

3) 通过随机方式产生  $n$  个  $L$  位二进制串作为初始群体的一个个体;

4) 不断产生初始个体,直到初始群体中个体数目为  $m$  为止

### (4) 群体繁殖

1) 选择 从上代群体中选择一定数量的个体,为参与下代群体繁殖的父代个体。选择个体的原则是使适应度大的个体被选择的概率亦大。常用的方法主要有轮盘选种法、RSIS 选种法、线性比例模型法等。

轮盘选种法是将所有个体的适应度之和看作一个轮盘,根据适应度的大小,将每个个体与轮盘中某些部分对应起来,然后旋转轮盘上的指针,指针所在处对应的个体就被选中。不断重复上述过程,直到选择到所需要的个体数目为止。

按照下述步骤就可利用计算机来实施轮盘选种法:

① 对上代群体中所有个体的适应度进行累加得适应度之和  $\sum F$ ;

② 根据各个体的适应度值的大小,将各个体与  $[0, \sum F]$  上的某区域建立对应关系;

③ 在  $[0, \sum F]$  范围内产生一个随机数;

④ 随机数所在的区域对应的个体被选择。

显然,个体的适应度值越大,在采用轮盘选种法时被选择的几率越大。当然,在实际随机选择时,有时适应度最大的个体也有可能不被选中,因而亦会出现群体退化的现象。

从上一代群体中选择得到的个体作为参与下一代群体繁殖的父代个体。

2) 产生后代个体 杂交、突变等基本繁殖算子是依靠随机的方式来选取的。一些学者研究表明,随机概率在 0.6~0.8 范围时利用杂交算子来产生后代个体,随机概率在 0.01~0.02 范围时采用突变算子来产生后代个体,对大多数优化问题比较合适。

利用选择的基本繁殖算子产生个体,直到产生的后代个体数目达到群体规模数。

### (5) 群体收敛判别

群体进化收敛性可通过各代群体平均适应度变化率和最优个体适应度变化率等指标来判别。如果群体平均适应度变化率和最优个体适应度变化率小于许可精度,则可以认为群体进化处于稳定状态,群体进化基本收敛,可结束群体进化过程,否则继续群体的进化过程。

### (6) 输出最优解

在群体中选择适应度值最大的个体,然后按式 (44.1-80) 对最优个体进行转化,就可得到优化问题的最优解和目标函数值。

对于约束优化问题,可先将约束优化问题转化为无约束优化问题,然后利用遗传算法来求解。

## 9 神经网络优化方法

人工神经网络是由大量的非线性处理单元,按照人脑的结构广泛互连而成,是对人脑结构和功能的高度抽象和模拟。人工神经网络与传统的 Von Neumann 结构计算机有着本质的区别,具有生物神经网络系统一些固有的特点。人工神经网络是以神经元间广泛的互连分布来存储信息,以非线性神经元来协同处理信息。尽管每个非线性处理单元的功能简单,但大量的处理单元并行作业时能实现强大的计算功能。自 1943 年美国心理学家 MC Culloch 和数学家 Pitts 提出形式神经元的数学模型(即 MP 模型)以来,神经网络得到了极大的发展和极为广泛的应用。神经网络为解决优化设计问题提供了新的手段。

### 9.1 神经网络优化方法基本思路

大家知道,依据优化问题设计变量的类型,可将优化设计问题分为连续变量优化设计问题、混合离散变量优化设计问题等。因此神经网络的优化方法包括连续变量、混合离散变量两方面优化问题的神经网络优化方法。

#### (1) 连续变量优化的神经网络优化方法思路

人工神经网络系统是以大规模、非线性、并行的方式来进行信息的处理。该神经网络系统趋向稳态的过程,就是神经网络系统的计算过程。在神经网络系统趋向稳态的过程中,其相应的能量函数单调递减。神经网络系统将沿非平衡轨道趋向平衡稳定状态。如果神经网络系统的平衡点是稳定的,那么这种稳定点同时也是渐近稳定点,并且这种稳定点是该神经网络系统的能量函数的一个局部极小点。因此,神经网络系统趋于稳态的过程,是在时间历程上向系统的能量函数单调减小方向运动的过程,也就是该系统能量函数向最优解逼近的过程。

先考虑无约束的连续变量优化问题。如果无约束连续变量优化问题的目标函数与神经网络系统的能量函数相对应,那么该连续变量优化问题从初始点向最优解逼近的过程,就可通过该神经网络系统从初始状态趋向稳态过程来实现。神经网络系统的初始态对应着无约束连续优化问题的初始解,神经网络系统的稳定态对应着无约束连续变量优化问题的最优解。这就是利用神经网络系统进行连续变量优化问题优化计算的主要思路。

对于有约束连续变量优化问题,我们采用罚函数的方法将有约束的优化问题转化为无约束连续变量优

化问题。然后再采用解决无约束连续变量优化问题的神经网络方法来求解。

#### (2) 混合离散变量优化的神经网络优化方法思路

与连续变量优化问题相比,混合离散变量优化问题的设计域有些特殊。可将混合离散变量优化问题看作是某些设计变量设计域为离散点列的连续变量优化设计问题。因而用神经网络解决混合离散变量优化问题时,可等同于在用神经网络解决连续变量优化问题同时考虑一些设计变量特殊的设计域。因此用神经网络解决混合离散变量优化问题的关键内容在于如何构造合适的能量函数及如何确保 Hopfield 神经网络系统稳定点停留在离散点设计域。构造一个合适的能量函数,可使优化问题求解对应于 Hopfield 神经网络系统稳定点的求解,从而为神经网络解决优化问题奠定基础。Hopfield 神经网络系统能否停留在混合离散优化问题设计变量的离散点域内,关系到能否满足优化问题的离散设计域特殊性要求、是否是离散变量优化问题的可行解。

由于连续 Hopfield 神经网络模型的本身特点,Hopfield 神经网络系统稳定点求解不能完全确保各神经单元最终收敛于优化问题设计变量的离散点域内。因此,为了保证用神经网络求解混合离散变量优化问题的有效性,必须对神经网络的计算过程进行调整。在神经网络的计算过程中,可采取学习策略使得神经单元收敛于优化问题设计变量的离散点域。在学习策略中,将各离散变量的离散点作为样本点输入到某种学习网络中,通过网络反复不断的学习,使在神经网络计算过程中如果神经元不能满足优化离散解的要求时,能进行样本点的记忆。可以采用 Hopfield 神经网络来实现上述功能。

### 9.2 神经网络优化方法

#### (1) 连续变量优化问题的神经网络优化方法

连续变量优化的神经网络优化方法包括如下两个主要步骤:

1) 建立优化问题的神经网络模型 根据连续变量优化问题选择合适的神经网络系统,将优化问题以神经网络系统来表达,建立无约束连续变量优化问题目标函数与神经网络系统能量函数一一对应的关系,使得无约束连续变量优化问题的求解过程等效转化为神经网络系统趋向稳定态的过程。建立优化问题的神经网络模型,亦可称为优化问题的神经网络模型化。

2) 优化问题神经网络模型的求解 求解连续变量优化问题的神经网络模型,可依据已构造的能量函数来构造合适的神经网络系统,然后选用数值计算方法对该神经网络系统进行稳定态的仿真计算。当然,也可

通过相应的数字或模拟电路或人工神经网络芯片来求解。

图 44.1-1 是无约束连续变量优化问题的神经网络优化方法的流程图。

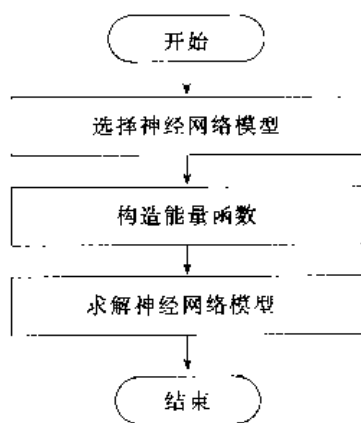


图 44.1-1 无约束优化神经网络优化方法流程图

对于有约束连续变量优化问题，可以采用罚函数的方法将有约束的优化问题转化为无约束连续变量优化问题；采用无约束连续变量优化问题的神经网络方法来求解。图 44.1-2 是有约束连续变量优化问题的神经网络优化方法的流程图。

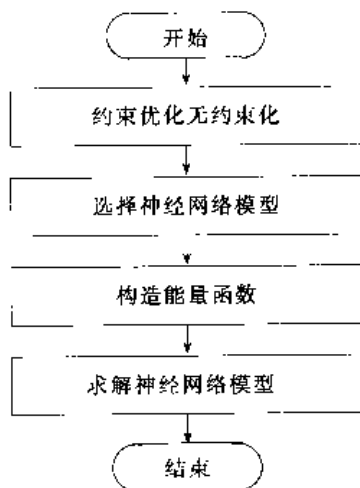


图 44.1-2 有约束优化神经网络优化方法流程图

## (2) 混合离散变量优化问题的神经网络优化方法

混合离散变量优化问题的神经网络优化方法包括如下两个主要步骤：

1) 建立优化问题的神经网络模型 根据优化问题的性质将优化目标函数与神经网络的能量函数联系

起来。将优化问题的设计变量对应于神经网络的单元状态，通过神经网络系统固有的能量函数的极小化来得到问题的最优解。对于有约束的离散变量优化问题，先将优化问题的约束条件以罚项的形式进行处理，然后再构造能量函数。

2) 优化问题神经网络模型的求解 由于工程中的离散变量优化设计问题大多表现为高度非线性的特性，使得混合离散变量优化问题存在更多的局部解。因此在混合离散变量优化问题求解过程中，跳出局部解区域进入全局最优解区域更显得十分重要。在优化问题的神经网络模型求解过程中，采用了模拟退火算法。模拟退火算法是一种能使优化问题获得全局最优解的方法。模拟退火算法借助固体物质退火处理的思路。如果在系统的状态变量变化后，能量函数有更低的值，那么就接受这种有益的变化；如果在系统的状态变量变化后，能量函数没有更低的值，那么就按预先确定概率方式来选择性地接受这种变化。模拟退火算法实际上给系统的变量引入一些“噪声”，使系统更有可能跳出能量函数的局部最优区域而向全局最优区域搜索。

混合离散变量优化问题的神经网络模型求解一般要先给定初始参数和初始点，然后以初始点为基础进行迭代计算；在迭代的计算过程中，需要考虑设计变量的离散要求。混合离散变量优化问题的神经网络模型求解过程如下：

1) 给定初始参数、所有神经元的初始输入值、系统的初始状态温度、模拟退火算法中性能变差可接受的次数等；

2) 给定初始迭代点；

3) 计算各神经元状态变化率；

4) 计算后续时间阶段各神经元状态，并计算出各设计变量的数值；将设计变量处理到所要求的离散设计点列；

5) 如果由步骤 4) 得到的设计新点使得原优化设计问题对应的能量函数减少（即目标函数值减小），则接受这种变化，并转步骤 7)；否则转步骤 3)；

6) 用模拟退火算法判断当前设计新点可接受与否；如可接受，转步骤 7)；否则转步骤 3)；

7) 判别搜索过程是否结束。如不结束，则转步骤 6)。

在混合离散优化的神经网络模型求解算法中，如果满足接受系统性能变差状态的次数超过一定值等条件，即可终止计算过程。在迭代过程中所得的最佳点，即为优化问题的最优解。

混合离散优化问题的神经网络求解算法流程如图 44.1-3 所示。

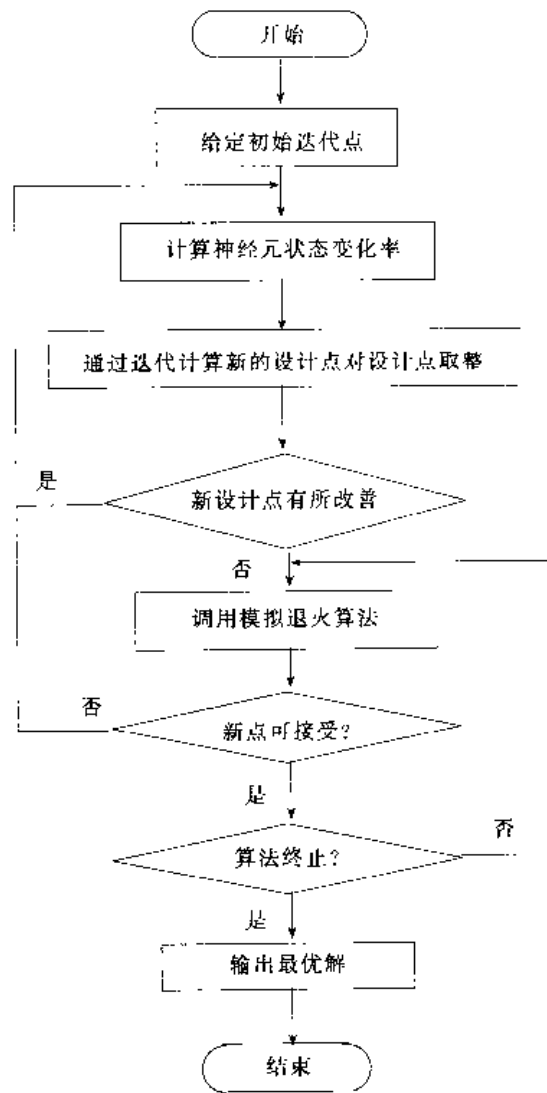


图 44.1-3 混合离散变量优化问题求解算法流程图

## 第2章 机构优化设计

### 1 机构优化设计概述

机构是具有确定运动的构件系统,用于运动的传递和变换,它是机械产品最基本的运动单元。机构设计是机械设计中基础的问题。常用的机构包括齿轮传动机构、凸轮机构、平面连杆机构、空间连杆机构等。

机构运动分析包括机构的位置、速度和加速度分析。建立机构从动件的输出参量、其他运动参量与机构结构参数之间的基本关系式,得到从动件输出变量与上动件输入变量以及机构结构参数之间位置或位移的数学方程式,这样就可通过建立的数学方程式来进行机构的位置分析。在进行机构位置分析基础上,通过计算机构位置或位移数学方程式对时间的一次导数,就可进行机构的速度分析;计算机构位置或位移数学方程式对时间的二次导数,就可进行机构的加速度分析。机构动力分析是指在给定机构运动和输出外力条件下进行输入扭矩和运动副反力的动态静力计算;在所有输入和输出外力给定条件下进行机构各刚性构件质量实际运动规律计算。机构动力分析十分复杂、涉及面广,主要内容有变质量机构动力学分析、振动结构动力学分析、刚性转子动力学分析等。通过机构的分析,可在已知机构尺寸参数基础上计算机构的运动轨迹,也可在已知机构预计运动轨迹基础上计算机构的尺寸参数。

机构设计基本问题包括实现预定运动的机构类型、构件联结形式及构件数目、机构的各种尺寸参数和物理参数等。机构设计的主要过程如下:

1) 根据产品的功能要求进行机构类型的选择。

2) 确定机构运动的形式。

3) 根据运动学和动力学方面的要求进行机构的运动学和动力学分析,确定机构的机构长度、支点位置等几何参数,质量质心位置等物理参数。

4) 进行机构的结构强度和工艺设计。

一般利用机构可以实现如下功能:

1) 刚体的导向功能,即可采用刚体的导向机构来引导刚体,使得机构通过空间一系列预定的位置。

2) 实现预定轨迹功能,即通过机构来引导刚体上的某一点,使得机构通过空间指定轨迹上的一系列点。

3) 实现预定函数功能,即设计的机构所产生的输出为输入运动的指定函数。

在进行机构设计时,同样类型的构件组合起来可以实现多种确定的运动,因此机构设计中存在构件组

合优选的问题。同样的构件在采用不同联结方式和形式时,构件也有着不同的运动学和动力学性能,因此机构设计时就存在机构性能优化的问题。从机构性能分析角度,可将机构优化设计问题分为二类:

1) 满足运动学要求建立目标函数的优化问题。这类优化问题主要解决有关再现给定的函数、轨迹、连杆位置等参量优化设计问题。

2) 满足动力学要求建立目标函数的优化问题。这类优化问题主要解决有关机构外力和质量方面的变量优化设计问题。

机构优化设计是机械设计中应用优化技术比较早和比较成功的领域之一。鉴于机构设计中存在着大量的优化问题,国内外许多学者在机构优化设计方面开展了大量的工作,先后对不同种类的机构优化问题进行了深入的研究。例如平面四杆机构、平面五杆机构、柔性连杆机构以及复杂的组合机构。并解决了大量实际工程应用中的难题。

在有关机构动力学优化问题方面,机构优化设计领域技术人员先后解决了丝织机投梭机构的动力学模型优化、柔性连杆机构的震动力优化、高速凸轮机构动态优化、凸轮从动杆机构动态优化、弹性连杆机构动力优化等问题。

近几年来,模拟生物遗传进化的遗传算法在机构优化设计方面也得到了应用。遗传算法先后被用来解决平面曲柄摇杆机构再现轨迹的优化设计、多杆压力机的工作机构优化、包装机组合推料机构优化设计、液压缸平衡机构的优化设计、平面连杆变幅机构轨迹综合优化、封闭行星齿轮机构优化、平面四杆机构轨迹优化综合、内燃机主副连杆机构优化设计、飞剪机剪切机构优化设计、引纬机构优化设计、冲压机构优化设计等问题。模糊优化、稳健优化、模糊可靠优化方法也被用来解决机构优化设计问题,例如凸轮机构的模糊优化设计问题、曲柄滑块机构的模糊可靠性优化设计问题和轨迹再现机构的模糊稳健优化设计问题。神经网络和混沌优化方法也在机构优化设计方面得到了应用,例如采用神经网络的装载机连杆机构优化设计、基于人工神经网络和优化技术的连杆机构优化、神经网络的游梁式抽油机四连杆机构的优化设计、齿轮传动机构人工神经网络辅助优化设计、采用混沌优化方法进行的平面四杆机构的函数发生器综合优化和平面连杆机构近似综合的混沌优化等。

## 2 连杆机构优化设计

连杆机构是最常用的机构,因此连杆机构优化设计在机构设计中十分重要,研究工作开展得也最为广泛。大量的文献介绍有关平面四杆机构、平面五杆机构、柔性连杆机构、曲柄连杆机构、槽轮连杆机构、凸轮连杆组合机构和齿轮连杆等机构的优化。鉴于四连杆机构的典型性,本节结合四连杆机构的函数再现优化设计问题,阐述连杆机构优化问题的一般方法及流程。

四连杆机构的优化设计就是对四连杆机构的参量进行优化调整,使得机构给定的运动和机构所实现的运动之间误差最小。因此四连杆机构的优化设计的过程,就是寻找使得四连杆机构运动误差最小的一组机构设计参量。四连杆机构设计参量确定后,就可认为实现了机构的优化设计。

四连杆机构的优化设计包括四连杆机构优化模型建立和优化模型求解二个主要过程。通过对四连杆机构的分析确定优化方案,确定设计变量,给出目标函数,并将机构设计约束条件,如杆长条件、传动角条件等,写成相应的约束条件,即可建立机构优化设计模型。

下面介绍四连杆机构函数再现优化设计模型的建立。连杆机构函数再现设计主要通过选取输入构件和输出构件相对应若干位置、采用机构图解法或分析法确定机构各参数。图 44.2-1 是典型的平面铰链四杆机构,  $l_1$ 、 $l_2$ 、 $l_3$  和  $l_4$  分别表示于四个构件的长度,杆  $AB$  是输入构件。假设图 44.2-1 所示的平面铰链四杆机构再现给定函数为  $\varphi_0 = f(\theta_0)$ , 即  $\varphi_0 = \varphi_1 - \varphi_2$ ,  $\theta_0 = \theta_1 - \theta_2$ , 则机构位置取决于  $l_1$ 、 $l_2$ 、 $l_3$ 、 $l_4$ 、铰链  $A$  的位置  $(a_x, a_y)^T$ 、 $AD$  与机架  $x$  轴夹角  $\beta$  以及输入构件转角  $\theta$  等七个变量。

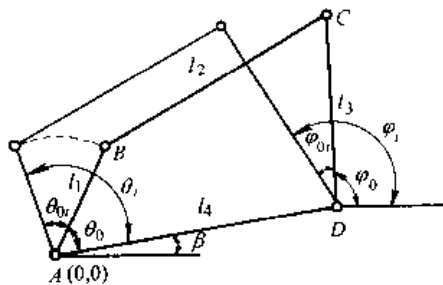


图 44.2-1 平面铰链四杆机构

为简化问题,可令  $A$  的位置为  $(0, 0)^T$ ,  $\beta = 0$ ,  $l_4$  构件的长度为 1 (参考构件), 由此可将问题维数降为四维,并不影响构件输入、输出的函数关系。由此可以得到输出构件转角  $\varphi_0$  与输入构件转角  $\theta_0$  之间的函数关系式:

$$\varphi_0 = f(l_1, l_2, l_3, \theta_0, \theta_0) \quad (44.2-1)$$

机构优化设计目标就是使得输出构件转角与给定值在  $\theta_0$ ,  $i=1, 2, \dots, s$  所有位置上的误差最小。因此机构优化设计的目标函数可用下式表示,

$$f(x) = \sum_{i=1}^s \omega_i [f(l_1, l_2, l_3, \theta_0, \theta_0) - \bar{f}(\theta_0)]^2 \quad (44.2-2)$$

当输入构件转角为  $\theta_0$  时,输出构件转角  $\varphi_0$  可由下式求得,

$$\varphi_0 = \beta_0 + \delta_0 - \beta_1 - \delta_1 \quad (44.2-3)$$

$$\text{式中 } \beta_0 = \arctan \frac{l_1 \sin \theta_0}{1 - l_1 \cos \theta_0};$$

$$\delta_0 = \arccos \frac{l_1^2 - l_2^2 + l_3^2 + 1 - 2l_1 \cos \theta_0}{2l_3 \sqrt{l_1^2 + 1 - 2l_1 \cos \theta_0}};$$

$$\beta_1 = \arctan \frac{l_1 \sin(\theta_0 + \theta_0)}{1 - l_1 \cos(\theta_0 + \theta_0)};$$

$$\delta_1 = \arccos \frac{l_1^2 - l_2^2 - l_3^2 + 1 - 2l_1 \cos(\theta_0 + \theta_0)}{2l_3 \sqrt{l_1^2 + 1 - 2l_1 \cos(\theta_0 + \theta_0)}}.$$

$$\text{所以 } \varphi_0 = \arctan \frac{l_1 \sin \theta_0}{1 - l_1 \cos \theta_0}$$

$$+ \arccos \frac{l_1^2 - l_2^2 + l_3^2 + 1 - 2l_1 \cos \theta_0}{2l_3 \sqrt{l_1^2 + 1 - 2l_1 \cos \theta_0}}$$

$$- \arctan \frac{l_1 \sin(\theta_0 + \theta_0)}{1 - l_1 \cos(\theta_0 + \theta_0)}$$

$$- \arccos \frac{l_1^2 - l_2^2 - l_3^2 + 1 - 2l_1 \cos(\theta_0 + \theta_0)}{2l_3 \sqrt{l_1^2 + 1 - 2l_1 \cos(\theta_0 + \theta_0)}},$$

$$i=1, \dots, s \quad (44.2-4)$$

将上式代入式 (44.2-3), 并令  $x = [x_1, x_2, x_3, x_4]^T$  代表设计变量  $l_1$ 、 $l_2$ 、 $l_3$  及  $\theta_0$ , 机构优化设计目标函数可写为

$$f(x) = \sum_{i=1}^s \omega_i \left[ \arctan \frac{x_1 \sin x_4}{1 - x_1 \cos x_4} \right. \\ + \arccos \frac{x_1^2 - x_2^2 + x_3^2 + 1 - 2x_1 \cos x_4}{2x_3 \sqrt{x_1^2 + 1 - 2x_1 \cos x_4}} \\ - \arctan \frac{x_1 \sin(x_4 + \theta_0)}{1 - x_1 \cos(x_4 + \theta_0)} \\ - \arccos \frac{x_1^2 - x_2^2 - x_3^2 + 1 - 2x_1 \cos(x_4 + \theta_0)}{2x_3 \sqrt{x_1^2 + 1 - 2x_1 \cos(x_4 + \theta_0)}} \\ \left. - f(\theta_0) \right]^2 \quad (44.2-5)$$

机构优化设计的约束条件应根据机构设计的实际情况确定。例如曲柄摇杆式四连杆机构必须满足如下关系式:

$$l_1 + l_2 \leq l_3 + l_4$$

$$\begin{aligned} l_1 + l_3 &\leq l_2 + l_4 \\ l_1 + l_4 &\leq l_2 + l_3 \end{aligned}$$

或

$$\begin{cases} 1 - l_1 - l_2 + l_3 \geq 0 \\ 1 - l_1 - l_2 - l_3 \geq 0 \\ -1 - l_1 + l_2 + l_3 \geq 0 \end{cases} \quad (44.2-6)$$

如果机构要求传动灵活可靠, 则传动角  $\gamma$  应满足

$$[\gamma]_{\min} \leq \gamma \leq [\gamma]_{\max}$$

或

$$\cos[\gamma]_{\max} \leq \cos\gamma \leq \cos[\gamma]_{\min}$$

其中

$$\cos\gamma = \frac{l_2^2 + l_3^2 - l_1^2 - 1 + 2l_1 \cos(\theta_0 + \theta_{0'})}{2l_2 l_3}$$

从上式可知, 传动角  $\gamma$  随  $\theta_0 + \theta_{0'}$  的变化而变化, 当  $\cos(\theta_0 + \theta_{0'})$  为最大值时,  $\gamma$  为最小,  $\cos(\theta_0 + \theta_{0'})$  为最小值时,  $\gamma$  为最大。要满足上式条件, 约束方程应为

$$\begin{cases} \cos[\gamma]_{\min} - \frac{l_2^2 + l_3^2 - l_1^2 - 1 + 2l_1 [\cos(\theta_0 + \theta_{0'})]_{\max}}{2l_2 l_3} \geq 0 \\ \frac{l_2^2 + l_3^2 - l_1^2 - 1 - 2l_1 [\cos(\theta_0 + \theta_{0'})]_{\min}}{2l_2 l_3} \leq \cos[\gamma]_{\max} \\ \frac{l_2^2 + l_3^2 - l_1^2 - 1 - 2l_1 [\cos(\theta_0 + \theta_{0'})]_{\min}}{2l_2 l_3} \geq 0 \end{cases}$$

曲柄摇杆机构有  $[\cos(\theta_0 + \theta_{0'})]_{\max} = 1$ ,  $[\cos(\theta_0 + \theta_{0'})]_{\min} = -1$ , 因此, 约束方程为

$$\begin{cases} \cos[\gamma]_{\min} - \frac{l_2^2 + l_3^2 - (1 - l_1)^2}{2l_2 l_3} \geq 0 \\ \frac{l_2^2 + l_3^2 - (1 + l_1)^2}{2l_2 l_3} - \cos[\gamma]_{\max} \geq 0 \end{cases} \quad (44.2-7)$$

当所选定的设计变量为构件长度时, 则构件长度必须是正数, 即约束方程为

$$\begin{cases} l_1 - \Delta l \geq 0 \\ l_2 - \Delta l \geq 0 \\ l_3 - \Delta l \geq 0 \end{cases}$$

式中  $\Delta l$  是为了使构件长度不小于  $\Delta l$  而设的。

此外, 由于具体结构尺寸的限制, 往往对某些构件的长度限定在某一范围内选取, 例如连杆 BC 的长度  $l_2$  最短为  $l_1$  的  $\lambda'$  倍, 最长为  $l_1$  的  $\lambda''$  倍, 即

$$\lambda' l_1 \leq l_2 \leq \lambda'' l_1$$

则约束方程为

$$\begin{cases} l_2 - \lambda' \geq 0 \\ \lambda'' - l_2 \geq 0 \end{cases} \quad (44.2-8)$$

下面介绍为再现函数为  $y = \sin(\pi x)$ ,  $0 \leq x \leq 1$  的曲柄摇杆机构的优化设计。

先变换给定函数  $y = \sin(\pi x)$  为  $\varphi_0 = \bar{f}(\theta_{0'})$ , 并设输入构件初始角为  $\theta_0$ , 输出构件初始角为  $\varphi_0$ , 选取输入构件的转角为  $180^\circ$ , 输出构件的转角为  $30^\circ$ 。当输入构件从  $\theta_0$  转到  $\theta_0 + 90^\circ$  时, 输出构件从  $\varphi_0$  转到  $\varphi_0 + 30^\circ$ , 输入构件从  $\theta_0 + 90^\circ$  转到  $\theta_0 + 180^\circ$  时, 输出构件则从  $\varphi_0 + 30^\circ$  转回到  $\varphi_0$ 。显然有,  $\frac{x}{180^\circ} = \frac{1}{180^\circ}$  及  $\frac{y}{30^\circ} = \frac{1}{30^\circ}$ , 即  $x = \frac{1}{180^\circ} \theta_{0'}$  及  $y = \frac{1}{30^\circ} \varphi_{0'}$ 。代入函数式  $y = \sin(\pi x)$ ,  $0 \leq x \leq 1$  得:

$$\varphi_{0'} = 30^\circ \sin\left[\frac{\pi}{180^\circ} \theta_{0'}\right], \quad 0^\circ \leq \theta_{0'} \leq 180^\circ$$

设将输入构件的转角  $180^\circ$  均分成 20 等分, 则  $\theta_{0'}$

$$= \frac{180^\circ}{20} i = 9^\circ i, \quad i = 1, \dots, 20$$

取权因子  $\omega_i = 1$ , 再令  $x = [x_1, x_2, x_3, x_4]^T$  代表设计变量  $l_1, l_2, l_3$  及  $\theta_{0'}$ , 则由式 (44.2-5) 得曲柄摇杆机构优化目标函数为

$$\begin{aligned} f(x) = & \sum_{i=1}^{20} \left[ \arctan \frac{x_1 \sin x_4}{1 - x_1 \cos x_4} \right. \\ & + \arccos \frac{x_1^2 - x_2^2 + x_3^2 + 1 - 2x_1 \cos x_4}{2x_2 \sqrt{x_1^2 + 1} - 2x_1 \cos x_4} \\ & - \arctan \frac{x_1 \sin(x_4 + \theta_{0'})}{1 - x_1 \cos(x_4 + \theta_{0'})} \\ & - \arccos \frac{x_1^2 - x_2^2 + x_3^2 + 1 - 2x_1 \cos(x_4 + \theta_{0'})}{2x_2 \sqrt{x_1^2 + 1} - 2x_1 \cos(x_4 + \theta_{0'})} \\ & \left. - \frac{\pi}{6} \sin\left(\frac{\pi}{180^\circ} \theta_{0'}\right) \right]^2 \end{aligned}$$

曲柄摇杆机构优化设计约束条件如下:

由式 (44.2-6) 得:

$$\begin{aligned} g_1(x) &= 1 - x_1 - x_2 + x_3 \geq 0 \\ g_2(x) &= 1 - x_1 + x_2 - x_3 \geq 0 \\ g_3(x) &= -1 - x_1 + x_2 + x_3 \geq 0 \end{aligned}$$

要求传动角满足  $30^\circ \leq \gamma \leq 135^\circ$ , 由式 (44.2-7) 得:

$$\begin{aligned} g_4(x) &= \frac{\sqrt{3}}{2} - \frac{x_2^2 + x_3^2 - (1 - x_1)^2}{2x_2 x_3} \geq 0 \\ g_5(x) &= \frac{x_2^2 + x_3^2 - (1 + x_1)^2}{2x_2 x_3} + \frac{1}{\sqrt{2}} \geq 0 \end{aligned}$$

根据机构结构尺寸, 要求各构件长度相对机架的尺寸在给定的范围内, 由式 (44.2-8) 得

$$\begin{aligned} g_6(x) &= x_1 - 0.1 \geq 0 \\ g_7(x) &= 0.5 - x_1 \geq 0 \\ g_8(x) &= x_2 - 0.1 \geq 0 \\ g_9(x) &= 1.4 - x_2 \geq 0 \\ g_{10}(x) &= x_3 - 0.1 \geq 0 \\ g_{11}(x) &= 0.9 - x_3 \geq 0 \end{aligned}$$



因此曲柄摇杆机构优化设计模型如下:

min.

$$f(x) = \sum_{i=1}^5 \left[ \arctan \frac{x_1 \sin x_4}{1 - x_1 \cos x_4} \right. \\ + \arccos \frac{x_1^2 - x_2^2 + x_3^2 + 1 - 2x_1 \cos x_4}{2x_1 \sqrt{x_1^2 + 1} - 2x_1 \cos x_4} \\ - \arctan \frac{x_1 \sin(x_4 + \theta_0)}{1 - x_1 \cos(x_4 + \theta_0)} \\ \left. - \arccos \frac{x_1^2 - x_2^2 + x_3^2 + 1 - 2x_1 \cos(x_4 + \theta_0)}{2x_1 \sqrt{x_1^2 + 1} - 2x_1 \cos(x_4 + \theta_0)} \right] \\ - \frac{\pi}{6} \sin \left[ \frac{\pi}{180} \theta_0 \right]^2$$

s. t.

$$g_1(x) = 1 - x_1 - x_2 + x_3 \geq 0$$

$$g_2(x) = 1 - x_1 + x_2 - x_3 \geq 0$$

$$g_3(x) = -1 - x_1 - x_2 + x_3 \geq 0$$

$$g_4(x) = \frac{\sqrt{3}}{2} - \frac{x_2^2 + x_3^2 - (1 - x_1)^2}{2x_2 x_3} \geq 0$$

$$g_5(x) = \frac{x_2^2 + x_3^2 - (1 + x_1)^2}{2x_2 x_3} + \frac{1}{\sqrt{2}} \geq 0$$

$$g_6(x) = x_1 - 0.1 \geq 0$$

$$g_7(x) = 0.5 - x_1 \geq 0$$

$$g_8(x) = x_2 - 0.1 \geq 0$$

$$g_9(x) = 1.4 - x_2 \geq 0$$

$$g_{10}(x) = x_3 - 0.1 \geq 0$$

$$g_{11}(x) = 0.9 - x_3 \geq 0$$

采用内点惩罚函数法和 POWELL 法来求解曲柄摇杆机构优化设计模型。选择初始惩罚参数  $\gamma_0 = 0.001$ , 递减函数  $e = 0.01$ , 初始点  $x^{(0)} = [l_1, l_2, l_3, \theta_0]^T = [0.3, 1, 0.6, 120^\circ]^T$ , 取惩罚函数法收敛精度  $\epsilon_1 = 0.01$ , POWELL 法目标函数值收敛精度  $\epsilon_2 = 10^{-4}$ , 一维搜索精度  $\epsilon = 10^{-3}$ , 解得最优解为

$$x^* = [0.1000011693, 1.011014285, 0.1856075558,$$

$$100.6264418^\circ]^T$$

$$f(x^*) = 0.0000716892864$$

### 3 凸轮机构优化设计

凸轮设计中存在着大量的优化问题, 国内外学者在凸轮机构方面开展了大量的优化设计工作, 例如平面凸轮机构、盘形凸轮机构、圆柱凸轮分度机构、齿轮凸轮组合机构、凸轮连杆组合机构、凸轮从动杆机构、分度凸轮机构、高速凸轮机构动态优化、凸轮从动杆机构动态优化、弹性连杆机构动力优化等问题。本节介绍内燃机凸轮机构的优化设计问题。

凸轮一般采用高次方多项式曲线作为基本工作段

气门升程曲线。升程曲线函数如下式所示:

$$h_x = c_0 + c_p \theta^p + c_q \theta^q - c_r \theta^r + c_s \theta^s + \dots \quad (4.1.2-9)$$

式中  $h_x$  —— 气门升程 (mm);

$c_0, c_p, c_q, c_r, c_s$  —— 待定系数;

$p, q, r, s$  —— 幂指数;

$\theta$  —— 在上升阶段为  $(\varphi_0 + \varphi)/\varphi_0$ , 在下降阶段为  $(\varphi - \varphi_0)/\varphi_0$ , 其中  $\varphi_0$  为凸轮半包角 ( $^\circ$ ), 为凸轮的转角 ( $^\circ$ )。

在内燃机凸轮设计中, 可通过选取多项式的项数和幂数来确定升程曲线。对于内燃机配气机构的凸轮, 选择  $p=2, q=2n, r=2n+2m, s=2n+4m$  ( $n$  和  $m$  为任意正整数,  $n=3, 4, 5, \dots; m=1, 2, 3, \dots$ ) 就能满足设计要求。内燃机配气机构设计中, 丰满系数是内燃机配气机构的重要性能指标, 因此可将丰满系数  $\xi$  最大作为内燃机凸轮优化设计目标。有关内燃机凸轮设计的最大正加速  $a_{\max} \leq [a_{\max}]$ 、最大负加速度  $a_{\min} \geq [a_{\min}]$ 、凸轮轮廓最小曲率半径  $R_{\min} \geq [R_{\min}]$  (方括号内的均表示允许值)、 $1 \leq n \leq 20$  和  $1 \leq m \leq 20$  等可作为约束条件。

内燃机配气机构凸轮优化设计模型如下:

设计变量为

$$x = \begin{bmatrix} n \\ m \end{bmatrix} = \begin{bmatrix} n \\ m \end{bmatrix}$$

目标函数为

$$f(x) = -\xi \quad (4.4.2-10)$$

约束条件为  $g_1(x) = x_1 - 3 \geq 0$

$$g_2(x) = 20 - x_1 \geq 0$$

$$g_3(x) = x_2 - 1 \geq 0$$

$$g_4(x) = 20 - x_2 \geq 0$$

$$g_5(x) = [a_{\max}] - a_{\max} \geq 0$$

$$g_6(x) = a_{\min} - [a_{\min}] \geq 0$$

式中

$$\xi = \frac{1}{c_0} \left( c_0 + \frac{c_p}{p+1} + \frac{c_q}{q+1} + \frac{c_r}{r+1} + \frac{c_s}{s+1} \right)$$

$$a_{\max} = \left( \frac{\omega}{\varphi_0} \right)^2 [2c_p + q(q-1)c_q x_0^{q-2}$$

$$+ r(r-1)c_r x_0^{r-2} + s(s-1)c_s x_0^{s-2}]$$

$$a_{\min} = 2c_p \left( \frac{\omega}{\varphi_0} \right)^2$$

$$R_{\min} = R_0 + c_0 + \frac{a_{\min}}{(\omega/57.29577951)^2}$$

$$c_0 = h_{\max}$$

$$c_p = \frac{-c_q r s + v(qr + rs + sq - q - r - s + 1)}{(q-p)(s-q)(p-q)}$$

$$c_v = \frac{-c_0 r s p + v(rs + sp + pr - r - s - p + 1)}{(r - q)(s - q)(p - q)}$$

$$c_r = \frac{-c_0 s p q + v(sp + pq + qs - s - p - q + 1)}{(s - r)(p - r)(q - r)}$$

$$c_s = \frac{-c_0 p q r + v(pq + qr + rp - p - q - r + 1)}{(p - s)(q - s)(r - s)}$$

$$x_0 = 2m \sqrt{\frac{q(q-1)(q-2)c_q}{s(s-1)(s-2)c_s}}$$

$$v = \frac{v_0 \varphi_0}{\omega}$$

$h_{\max}$ 和  $v_0$  均为给定值。

以 12150L 型柴油机配气机构的双圆弧凸轮改进为高次多项式升程凸轮设计为例, 原设计双圆弧凸轮的几何参数最小圆弧曲率半径  $R_1 = 4.5\text{mm}$ , 圆弧曲率半径  $R_2 = 154.9\text{mm}$ , 基圆半径  $R_0 = 20\text{mm}$ , 最大升程  $h_{\max} = 13\text{mm}$ , 半包角  $\varphi_0 = 62^\circ$ ; 凸轮的转速  $n = 1000\text{r/min}$ , 角速度  $\omega = 6000^\circ/\text{s}$ , 气门落座的最大速度为  $0.983\text{m/s}$ , 最大正加速  $a_{\max} = 1479\text{m/s}^2$ , 最大负加速度  $a_{\min} = -312.54\text{m/s}^2$ 。改进设计时,  $R_0$ 、 $h_{\max}$ 、 $\varphi_0$  和  $\omega$  均保持不变; 取  $v_0 = 0.5\text{m/s}$ ; 要求  $a_{\max}$  和  $a_{\min}$  都不超出原凸轮的数值, 即  $[a_{\max}] = 1479.34\text{m/s}^2$ ,  $[a_{\min}] = -312.54\text{m/s}^2$ , 最小轮廓曲率半径不小于原凸轮数值, 即  $[R_{\min}] = 4.5\text{mm}$ 。

将上述参数代入式 (44.2-10) 描述的优化模型, 利用复合形法来求解内燃机配气机构凸轮优化设计模型。设计变量的边界约束为  $3 \leq x_1 \leq 20$ ,  $1 \leq x_2 \leq 20$ 。优化设计得到的最优解为  $x_1^* = n = 8$ ,  $x_2^* = m = 8$ 。与优化解相关的内燃机配气机构凸轮升程曲线参数为  $p = 2$ ,  $q = 16$ ,  $r = 32$ ,  $s = 48$ ,  $c_p = -15.8080898032$ ,  $c_q = 4.40734242818$ ,  $c_r = -2.040950520$ ,  $c_s = 0.442588032$ 。内燃机配气机构凸轮丰满系数、最大正加速度、最大负加速度、最小曲率半径与原双圆弧凸轮有关这方面的数值列于表 44.2-1。

表 44.2-1 优化结果

主要性能参数	高次方多项式凸轮	双圆弧凸轮
丰满系数 $\xi$	0.6105215978	0.6253
最大正加速度 $a_{\max}/(\text{m/s}^2)$	1468.900714	1479.340
最大负加速度 $a_{\min}/(\text{m/s}^2)$	-296.1099322	-312.540
最小曲率半径 $R_{\min}/\text{mm}$	6.0	4.5

## 4 机构多目标优化设计

在机构设计时, 会经常遇到多个性能指标都要求很高的情况。因此在进行机构的优化设计过程中, 应将这些性能指标作为设计的目标, 建立包含多个目标函数的综合优化模型。包含多个目标函数的优化模型被

称为多目标优化问题。一些学者应用多目标优化方法解决了滑轮组变幅机构多目标设计问题、门座起重机变幅机构的多目标优化问题和液压支架总体参数多目标优化设计问题。

多目标优化问题中目标函数之间关系十分复杂。有时目标函数间常常相互冲突, 而且许多目标是互补的, 例如机构的性能可靠性和费用是典型的相互冲突、互补的目标函数。在多目标优化问题中, 希望找到一组解  $(x_1, \dots, x_n)$ , 能够使一组目标函数  $f = (f_1, \dots, f_m)$  在某一可行域  $\Omega$  里极小或极大。与单目标优化问题不同, 要求各种目标函数同时到达极大或极小的解是很可能不存在。目前多目标优化方法包括加权方法、极大极小方法等。所谓加权方法, 就是将  $n$  个目标函数  $f_1, \dots, f_n$  按用户定义的加权系数  $w_1, \dots, w_n$  加权, 来获得一个目标函数。这种方法最为简单, 也最为通用。极大极小方法就是指极小化  $n$  个目标函数  $f_1, \dots, f_n$  的极大值, 将  $n$  个对应期望的目标值  $g_1, \dots, g_n$  转化为极小化加权意义下的最大偏差值。极大极小方法通常需要对期望目标值  $g_k$  多次选取, 对算法多次运行才能获得较满意的最优解。

针对多目标优化问题的复杂性, 人们提出了 Pareto 最优解的概念。假设  $x^1, x^2 \in \Omega$ ,  $x^1$  被  $x^2$  占优是指对部分  $i$ , 有  $f_i(x^1) \geq f_i(x^2)$ , 而对其他的  $j \neq i$ ,  $f_j(x^1) > f_j(x^2)$ 。 $x^0$  被称为 Pareto 最优解, 如果在  $\Omega$  中不存在任何  $x$  占优于  $x^0$ 。对于多目标优化问题, 如果不可能找到绝对最优解, 往往希望能找到问题的 Pareto 最优解, Pareto 最优解往往亦称为相对最优解, 亦指在不降低其他性能指标的前提下, 再也不能提高该性能指标。

目前在机构多目标优化设计方面, 人们常采用将加权因子把各目标函数结合成综合目标函数的多目标函数处理方法。多目标优化设计问题目前仍然是优化设计领域有待进一步研究的问题。因此在建立多目标优化模型时, 原则上应尽量控制目标函数的数目, 使同时求解的优化目标尽可能少些。

本节结合支撑掩护式液压支架总体设计实例介绍机构的多目标优化设计方法。

液压支架机构多目标优化设计模型共含 5 个目标函数、21 个优化设计变量、9 类 71 个约束条件。液压支架机构如图 44.2-2 所示。

在液压支架机构多目标优化设计模型中, 设计变量  $x_1 \sim x_8$  和  $x_{17}$  为四连杆机构掩护梁的结构参数;  $x_9 \sim x_{12}$  为底座结构参数;  $x_{13} \sim x_{16}$  为顶梁结构参数;  $x_{18}$  和  $x_{19}$  为后连杆倾角  $\alpha$  的下界和上界;  $x_{20}$  和  $x_{21}$  为前后立柱的工作阻力。设计参数符号的意义如表 44.2-2 所示。

在液压支架机构多目标优化设计模型中，从液压支架重量最轻等 5 个方面建立目标函数。

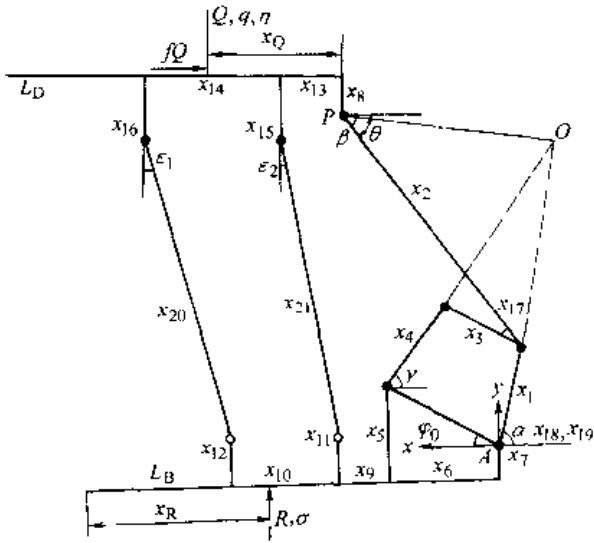


图 44.2-2 液压支架机构示意图

表 44.2-2 设计参数意义

符号	意义
$\alpha$	后连杆倾角
$\beta$	掩护梁倾角
$\gamma$	前连杆倾角
$\theta$	运动瞬时 O 压力角
$\epsilon_1$	前立柱倾角
$\epsilon_2$	后立柱倾角 $F$
$P$	顶梁和掩护梁铰点 $(x_p, y_p)$
$O$	运动瞬时 $(x_0, y_0)$
$Q, x_Q$	顶梁负荷与作用位置
$R, x_R$	底座反力与作用位置
$q, \eta$	平均支护强度与支护效率
$\sigma, \sigma_n$	最大比压与底座平均比压
$L_D$	顶梁和前立柱铰点到前端的距离
$L_B$	底座和前立柱铰点到前端的距离
$N_p$	等分后连杆倾角 $\alpha$ 的计算点数

1) 液压支架重量最轻设计目标

$$\min f_1 = G(x) \quad (44.2-11)$$

2) 梁端点水平偏移量加权平均和最小目标，即

$$\min f_2 = \sum_{i=1}^{N_p} \omega_{\sigma} (x_{p_i} - x_{p, N_p}) \quad (44.2-12)$$

式中  $\omega_{\sigma}$ ——加权因子。调整这组加权因子可以控制梁端点轨迹的形态，使之趋于更加合理。通过反复试算，可以确定合理的加权因子。

3) 支护效率加权平均和最大目标

$$\text{即 } \min f_3 = \sum_{i=1}^{N_p} \omega_{\eta} \eta_i \quad (44.2-13)$$

式中  $\omega_{\eta}$ ——加权因子。调整这些加权因子可以保证支架工作重要区段的支护效率比较高。

4) 底座比压加权平均和最小目标，即

$$\min f_4 = \sum_{i=1}^{N_p} \omega_{\sigma} \sigma_i \quad (44.2-14)$$

式中  $\omega_{\sigma}$ ——加权因子。调整这组加权因子可以抑制底座比压的尖峰值，使液压支架工作过程中的底座比压趋于平衡。

5) 掩护梁最大水平投影长度最小目标

$$\min f_5 = \max \{x_2 \cdot \cos \beta_i\} \quad (i = 1, 2, 3, \dots, N_p) \quad (44.2-15)$$

$\max \{x_2 \cdot \cos \beta_i\}$  是数列  $\{x_2 \cdot \cos \beta_i\}$  ( $i = 1, 2, 3, \dots, N_p$ ) 中的最大值，称为掩护梁最大水平投影长度。它对掩护梁乃至液压支架各构件的载荷都有相当大的影响。

在液压支架机构多目标优化设计模型中，约束条件包含如下 9 类 71 个约束：

1) 设计变量的边界约束条件为

$$\left. \begin{aligned} g_i(x) &= x_i - B_u \leq 0 \quad (i = 1, 2, \dots, 21) \\ g_i(x) &= B_L - x_i \leq 0 \quad (i = 22, \dots, 42) \end{aligned} \right\} \quad (44.2-16)$$

式中  $B_u$ ——设计变量  $x_i$  的上界；

$B_L$ ——设计变量  $x_i$  的下界。

$B_u, B_L$  的值应根据设计要求和经验取定。

2) 顶梁和底座长度的约束条件为

$$\left. \begin{aligned} g_{43}(\bar{x}) &= (L_D + x_{13} + x_{14}) - L_{DM} \leq 0 \\ g_{44}(\bar{x}) &= L_{D1} - (L_D + x_{13} + x_{14}) \leq 0 \\ g_{45}(\bar{x}) &= (L_B + x_6 + x_8 + x_{10}) - L_{BM} \leq 0 \\ g_{46}(x) &= L_{B1} - (L_B + x_6 + x_8 + x_{10}) \leq 0 \end{aligned} \right\} \quad (44.2-17)$$

式中  $L_{DM}, L_{D1}$ ——顶梁合理长度的极限；

$L_{BM}, L_{B1}$ ——底座合理长度的极限。

3) 双摇杆机构存在的条件为

$$\left. \begin{aligned} g_{47}(\bar{x}) &= (x_3 + x_4) - (L_Z + x_1) \leq 0 \\ g_{48}(\bar{x}) &= (x_1 + x_3) - (L_Z + x_4) \leq 0 \end{aligned} \right\} \quad (44.2-18)$$

$$\text{式中 } L_Z = \sqrt{x_5^2 + (x_6 - x_7)^2}.$$

4) 后连杆倾角的约束条件为

$$\left. \begin{aligned} g_{49}(\bar{x}) &= \alpha_{\min} - x_{18} \leq 0 \\ g_{50}(\bar{x}) &= x_{19} - \alpha_{\max} \leq 0 \end{aligned} \right\} \quad (44.2-19)$$

后连杆倾角的允许范围  $x_{18}$  和  $x_{19}$  主要是由经验确定的，而后连杆倾角的实际范围决定于其他几个设计变量的值。

$$\alpha_{\min} = \pi - \varphi_0 - \arccos \left[ \frac{L_Z^2 + x_1^2 - (x_3 + x_4)^2}{2L_Z x_1} \right]$$

$$\alpha_{\max} = \pi - \varphi_0 - \arccos \left[ \frac{L_2^2 + (x_1 - x_3)^2 - x_4^2}{2L_2(x_1 + x_3)} \right]$$

式中  $\varphi_0 = \arctan \left( \frac{x_3 - x_1}{x_4} \right)$ 。

5) 支架最低位置的约束条件为

$$\left. \begin{aligned} g_{51}(\bar{x}) &= (y_p + x_1 + x_3) - H_{\min} \leq 0 \\ g_{52}(\bar{x}) &= \beta_{\min} - \beta \leq 0 \\ g_{53}(\bar{x}) &= \gamma_{\min} - \gamma \leq 0 \\ g_{54}(\bar{x}) &= |\epsilon_1| - \epsilon_{1L} \leq 0 \\ g_{55}(\bar{x}) &= \epsilon_2 - \epsilon_{2L} \leq 0 \end{aligned} \right\} \quad (44.2-20)$$

6) 支架最高位置的约束条件为

$$\left. \begin{aligned} g_{56}(\bar{x}) &= H_{\max} - (y_p + x_1 + x_3) \leq 0 \\ g_{57}(\bar{x}) &= \beta - \beta_{\max} \leq 0 \\ g_{58}(\bar{x}) &= \gamma - \gamma_{\max} \leq 0 \\ g_{59}(\bar{x}) &= |\epsilon_1| - \epsilon_{1H} \leq 0 \\ g_{60}(\bar{x}) &= |\epsilon_2| - \epsilon_{2H} \leq 0 \end{aligned} \right\} \quad (44.2-21)$$

式中  $H_{\min}, H_{\max}$  — 液压支架设计高度最小、最大值；

$\beta_{\min}, \beta_{\max}$  — 掩护梁倾角允许最小、最大值；

$\gamma_{\min}, \gamma_{\max}$  — 前连杆倾角允许最小、最大值；

$\epsilon_{1L}, \epsilon_{2L}$  — 前、后立柱最大倾角限制；

$\epsilon_{1H}, \epsilon_{2H}$  — 前、后立柱最小倾角限制。

7) 运动参数约束为

$$\left. \begin{aligned} g_{61}(\bar{x}) &= |x_{p\max} - x_{p\min}| - \Delta_{\max} \leq 0 \\ g_{62}(\bar{x}) &= |\theta| \leq \theta_m \end{aligned} \right\} \quad (44.2-22)$$

式中  $x_{p\min}, x_{p\max}$  — 坐标  $x_p$  的极限值；

$\theta_m$  — 瞬行压力角的允许值；

$\Delta_{\max}$  — 梁端距的允许偏移量。

8) 支护性能参数约束条件为

$$\left. \begin{aligned} g_{63}(\bar{x}) &= [q] - q_r \leq 0 \\ g_{64}(\bar{x}) &= [\eta] - \eta_a \leq 0 \\ g_{65}(\bar{x}) &= \sigma_a - [\sigma] \leq 0 \\ g_{66}(\bar{x}) &= \sigma_{\max} - \sigma_b \leq 0 \end{aligned} \right\} \quad (44.2-23)$$

式中  $q_r, \eta_a, \sigma_a$  — 在支架高度变化范围内支护强度、支护效率和底座比压的平均值；

$[q], [\eta], [\sigma]$  — 支护强度、支护效率和底座比压的要求值或允许值；

$\sigma_{\max}$  — 底座比压的最大值；

$\sigma_b$  — 底座的抗压强度。

9) 力学参数约束条件为

$$\left. \begin{aligned} g_{67}(\bar{x}) &= M_1 - R_{10}M_{10} \leq 0 \\ g_{68}(\bar{x}) &= M_2 - R_{20}M_{20} \leq 0 \\ g_{69}(\bar{x}) &= M_3 - R_{30}M_{30} \leq 0 \\ g_{70}(\bar{x}) &= R_F - R_{40}R_{F0} \leq 0 \\ g_{71}(\bar{x}) &= R_B - R_{50}R_{B0} \leq 0 \end{aligned} \right\} \quad (44.2-24)$$

式中  $M_1, M_2, M_3$  — 可行设计的顶梁、掩护梁和底座的最大弯矩；

$M_{10}, M_{20}, M_{30}$  — 初始可行设计的顶梁、掩护梁和底座的最大弯矩；

$R_F, R_B, R_{F0}, R_{B0}$  — 可行设计的前、后连杆的最大载荷和初始可行设计的前、后连杆的最大载荷；

$R_{10}, R_{20}, R_{30}, R_{40}, R_{50}$  — 容许放大系数，均大于 1，根据设计意图决定。

上述约束条件中有关运动学参数、支护性能及力学参数的计算，可由运动学和力学分析计算公式求出。

因此液压支架机构多目标优化设计模型如下式所示：

Min.

$$\begin{aligned} f_1 &= G(\bar{x}) \\ f_2 &= \sum_{i=1}^{N_p} \omega_{xi} (x_{pi} - x_{p,N_p}) \end{aligned} \quad (44.2-25)$$

$$f_3 = \sum_{i=1}^{N_p} \omega_{\eta_i} \eta_i$$

$$f_4 = \sum_{i=1}^{N_p} \omega_{\sigma_i} \sigma_i$$

$$f_5 = \max |x_2 \cdot \cos \beta|$$

$$\bar{x} = [x_1, x_2, \dots, x_{21}]$$

$$\text{s. t. } g_u(\bar{x}) \leq 0 \quad (u = 1, 2, \dots, 71)$$

由于液压支架机构多目标优化设计模型包含 21 个设计变量、71 个约束条件和 5 个目标函数，优化模型十分复杂。另一方面约束条件存在着层次递进关系，例如如果液压支架结构高度约束不满足，后面的运动学、力学约束就无计算判别的必要。在求解液压支架机构多目标优化设计模型时，采用加权处理策略，构造综合目标函数

$$f(\bar{x}) = \sum_{j=1}^5 \frac{\omega_{f_j} f_j}{f_{j0}} \quad (44.2-26)$$

式中  $\omega_{f_j}, f_{j0}$  — 目标函数  $f_j$  的加权因子和初始可行设计的对应值。

另外通过检验约束条件和计算综合目标函数可以同步进行，可行性约束条件判别结束时，目标函数值的计算亦完成；预先可以将各类约束条件排出层次优先级，若前面的约束条件不满足，后面的约束条件就不用进行检验，从而加快收敛速度。优化设计有关参数和结

果如表 44.2-3 和表 44.2-4 所示。

表 44.2-3 优化前后设计变量值

设计参数	优化前数值	优化设计值	下界	上界
$x_1$	1850	1852.22	1600	2000
$x_2$	2400	2413.16	2000	2600
$x_3$	450	454.52	300	500
$x_4$	1850	1846.87	1600	2000
$x_5$	830	833.81	500	900
$x_6$	700	697.76	500	800
$x_7$	280	289.48	0	400
$x_8$	115	120.05	0	300
$x_9$	350	354.03	0	400
$x_{10}$	850	851.42	600	900
$x_{11}$	360	356.34	0	400
$x_{12}$	360	368.47	0	400

(续)

设计参数	优化前数值	优化设计值	下界	上界
$x_{13}$	300	300.66	100	400
$x_{14}$	1050	1050.93	800	1200
$x_{15}$	381	381.79	300	500
$x_{16}$	395	396.24	300	500
$x_{17}$	0°	0.05°	0°	10°
$x_{18}$	39°	39.335°	30°	50°
$x_{19}$	77°	76.817°	60°	90°
$x_{20}$	2352	2358.46	2000	2500
$x_{21}$	2352	2358.15	2000	2500

表 44.2-4 优化前后目标函数值

目标函数	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
优化前数值	9252.2	3021.17	92.49	121.12	2325.26
优化后数值	8740.96	1710.69	91.61	120.13	2335.21

## 第3章 机械零件优化设计

### 1 机械零件优化设计概述

机械零件是指直接加工而不经装配的机器组成单元。机械零件是机械产品或系统的基础,机械产品由若干零件和部件组成。按照零件的应用范围,可将零件分为通用零件和专用零件二类。通用的机械零件包括齿轮、弹簧、轴、滚动轴承、滑动轴承、联轴器、离合器等。

机械零件设计就是确定零件的材料、结构和尺寸参数,使零件满足有关设计和性能方面的要求。机械零件除一般要满足强度、刚度、寿命、稳定性、公差等级等方面的设计性能要求,还要满足材料成本、加工费用等方面的经济性要求。

常规的机械零件设计方法包括理论设计、经验设计和模型实验设计等方法。所谓理论设计,就是根据现有的设计理论和实验数据进行机械零件的设计。一般可根据零件的工作状况、工作要求,先进行零件的失效形式分析,确定设计零件适用的工作能力准则;然后按机械零件理论设计公式计算零件的形状和尺寸等方面的参数。机械零件的设计也可先参照现有零件的实物、图样或经验数据初步确定零件结构和尺寸参数;然后再依据工作能力准则有关的理论校核公式进行校核计算;如果初步确定的零件结构和尺寸参数不能通过校核计算,就要调整零件结构和尺寸参数,重新进行机械零件的校核计算。所谓经验设计,就是根据现有同类零件长期设计经验积累而得到的经验公式进行零件设计,或者依据现有同类零件而进行的类比设计。利用经验设计进行机械零件设计,设计方法简单方便;因此对于工作环境和要求相近、零件外形和结构已标准化的机械零件,经验设计法的设计效率较高,比较实用可行。目前通用零件例如齿轮等传动零件的设计多采用经验设计法。所谓模型实验设计,就是对机械零件采用模型实验的手段确定零件参数和结构,从而进行零件的设计。采用模型实验设计法的零件一般难以采用理论计算方法进行设计。采用模型实验设计,一般先将初步设计的零件按照一定的比例加工制作成较小的模型,然后通过模型实验对零件的性能进行测试和实验,最后再根据测试和实验结果进行零件设计方案的修改设计,从而完成零件的设计工作。

进行机械零件的设计,一般需要确定零件的计算载荷、计算准则及零件尺寸参数。零件计算载荷和计算准则的确定,应当依据机械产品的总体设计方案对零

件的工作要求进行载荷等方面的详细分析,在此基础上建立零件的力学模型,考虑影响载荷的各项因素和必要的安全系数,确定零件的计算载荷;对零件工作过程可能出现的失效形式进行分析,确定零件设计或校核计算准则。零件材料和参数的确定,应当依据零件的工作性质和要求,选择适合于零件工作状况的材料;分析零件的应力或变形,根据有关计算准则,计算确定零件的主要尺寸参数,并进行参数的标准化。

所谓机械零件优化设计是将零件设计问题描述为数学优化模型,采用优化方法求解一组零件设计参数。机械零件设计中包含了许多优化问题,例如零件设计方案的优选问题、零件尺寸参数优化问题、零件设计性能优化问题等。国内机械设计领域技术人员针对齿轮、弹簧、滚动轴承、滑动轴承、联轴器、离合器等零件优化设计问题开展了大量的工作,解决了齿轮传动比优化分配、各种齿轮参数优化、各种齿轮减速器优化设计、各种齿轮传动的可靠性优化、齿轮传动和减速器多目标优化设计、各种弹簧的优化设计、各种滚动轴承和滑动轴承的优化设计、各种联轴器和离合器的优化设计问题。

在齿轮传动方面,国内学者已经解决了双曲柄行星齿轮传动比的优化分配、二级行星齿轮减速器传动比优化分配、二级圆柱齿轮变速器齿数比优化分配、四级圆柱齿轮减速器传动比的优化分配、双圆弧齿轮传动优化设计、喷气织机卷取机构变换齿轮优化选配、曲柄式渐开线行星齿轮传动优化、NGW型行星齿轮传动优化、双级蜗杆减速器传动比分配优化设计等问题。国内有关学者还针对高速铁路机车牵引齿轮、大重合度圆柱斜齿轮、多级齿轮参数、仪表圆弧齿轮、轧机人字齿轮承载能力、齿轮传动的等弯曲强度、低速大输出扭矩齿轮传动啮合参数、塑料谐波齿轮传动、少齿差行星齿轮传动参数、跑合性能的双圆弧齿轮齿形参数、齿轮抗磨损、齿轮变位系数、压辊齿轮变位系数、渐开线齿轮齿根过渡曲线、采煤机摇臂传动齿轮变位系数、NGW型行星齿轮传动角变位参数、渐开线圆柱齿轮变位系数、降低斜齿噪声的齿轮修形参数等设计问题进行了优化设计。在齿轮减速器方面,已经开展了多级平行轴圆柱齿轮减速器、斜齿轮减速器、标准斜齿圆柱齿轮二级展开式减速器、三级齿轮传动减速器、单级圆柱齿轮减速器、双圆弧齿轮减速器、级间等强度圆锥圆柱齿轮减速器、双自由度锥面包络圆柱蜗杆传动系统等优化设计。

在机械零件弹簧方面,已经开展了压缩弹簧、碟形弹簧、波形弹簧、板弹簧、片弹簧等的优化设计工作,并针对弹簧用途和种类多的特点,专门开展了矿井提升机盘式制动器碟形弹簧、螺旋形压缩弹簧、受振载荷圆柱螺旋压缩弹簧、汽车悬架系统弹簧、变截面非对称钢板弹簧、圆柱螺旋压缩弹簧、调节弹簧、离合器拉式膜片弹簧、离合器碟形弹簧、矩形截面圆柱螺旋压缩弹簧、拖拉机上离合器弹簧、减振弹簧、实用气阀弹簧、波形弹簧、发动机气门弹簧、抽油泵圆柱弹簧、电器弹簧、非标准碟形弹簧、活塞式压缩气阀弹簧、锅炉安全阀弹簧、溢流阀调节弹簧等的优化设计。

在机械零件轴承方面,先后开展了滚动轴承额定寿命的优化选型、球面轴承结构参数的优化、液体动压润滑轴承的优化设计、传动轴承结构参数的优化设计、螺旋油楔动压轴承的优化设计、空气静压轴承的优化设计、调心滚子轴承的优化设计等方面的工作。

在机械零件联轴器和离合器方面,先后开展了圆柱螺旋弹簧联轴器的离散优化设计、挠性膜片联轴器优化设计、等刚度直杆弹簧联轴器的离散优化设计、轴向放置叠片弹簧联轴器的离散变量优化设计、直线齿廓蛇形弹簧联轴器优化设计、齿式联轴器齿形优化设计、无螺栓十字轴万向联轴器结构形状优化设计、电磁摩擦离合器的参数优化设计、多片式摩擦离合器优化设计、空气离心式离合器优化设计、重载机械膜片弹簧离合器优化设计、汽车离合器摩擦片数量选择及其参数优化设计、纺织机械离合器联结方式优化等工作。

模糊优化和可靠性优化方法可以应用于机械零件的优化设计。在齿轮传动方面,开展了齿轮传动的可靠性优化设计计算、行星齿轮传动的可靠性优化设计、行星齿轮减速器可靠性优化设计、行星齿轮传动的模糊优化设计、行星齿轮减速器的模糊多目标优化设计、斜齿圆柱齿轮减速器的概率优化设计、单级圆柱齿轮减速器的概率优化设计、圆柱齿轮啮合参数模糊优化设计、具有可靠度约束的齿轮传动的优化设计、具有可靠度约束的行星齿轮传动稳健优化设计、齿轮变位系数选择的模糊优化设计、齿轮传动最小体积的多目标模糊优化设计、基于模糊综合评判的斜齿轮减速器多目标模糊优化设计、应用灰色聚类的钻床齿轮变速箱优化设计、三级圆锥和圆柱齿轮减速器传动比最优分配多目标模糊优化设计、微线段齿轮滚刀齿廓曲线的多目标优化设计、圆锥齿轮传动模糊优化设计、圆柱齿轮传动的模糊优化设计、直齿圆柱齿轮的模糊优化设计、单级2KH行星齿轮装置的模糊优化设计、NGW型行星齿轮减速器的模糊优化设计、锥齿轮传动的模糊可靠性优化设计、双圆弧齿轮传动的模糊可靠性优化设计、齿轮减速器的模糊可靠性优化设计、具有动态约束

的齿轮传动模糊优化设计、基于模糊可靠度约束的准双曲面齿轮优化设计、斜齿圆柱齿轮传动的多目标模糊优化设计、复杂工况下齿轮接触强度的可靠性优化设计、2KH行星齿轮减速器的可靠性优化设计、圆柱蜗杆减速器的可靠性优化设计等工作。针对弹簧、轴承、离合器等零件,开展了钻采机械碟形弹簧可靠性优化设计、碟形弹簧的模糊可靠性优化设计、汽车离合器膜片弹簧的模糊可靠性优化设计、内燃机气门弹簧的模糊优化设计、承受交变载荷的碟形弹簧的可靠性优化设计、圆柱螺旋压缩弹簧的模糊优化设计和模糊可靠性优化设计、发动机气门弹簧的可靠性优化设计、内燃机气门压缩弹簧模糊优化设计、汽车离合器膜片弹簧的稳健优化设计、铁道车辆螺旋支承弹簧的可靠性优化设计、基于模糊可靠性分析的碟形弹簧优化、圆柱形扭转弹簧的可靠性优化和模糊可靠优化设计、螺旋弹簧可靠性多目标优化设计、膜片弹簧设计的概率优化、拉式膜片弹簧结构参数的模糊优化设计、基于模糊概率的溢流阀调压弹簧的优化设计、变截面钢板弹簧模糊优化设计、扩张式自激超越弹簧离合器强度可靠性优化设计、车辆用板弹簧的可靠性优化设计、双卷螺旋弹簧模糊可靠性优化设计、滑动轴承最小油膜厚度的模糊优化设计、流体动压润滑向心滑动轴承的可靠性优化设计、四旋盖真空封瓶机双向联轴器的模糊优化设计、牙嵌离合器的可靠性遗传优化设计、模块式单向超越离合器结构参数的模糊优化设计、扩张式自激超越弹簧离合器强度可靠性优化设计、干式摩擦离合器的模糊优化设计、片式摩擦离合器的模糊优化设计、汽车离合器的模糊优化设计、双频发电机组单向自动离合器工作可靠性的优化设计。

机械零件的多目标优化设计方面如齿轮传动、弹簧、离合器、轴承等,许多研究者研究了多目标、多参数的变速器齿轮系优化设计、谐波齿轮传动多目标优化设计、普通圆柱蜗杆传动的多目标模糊优化设计、蜗杆减速机的多目标优化设计、弧面蜗杆传动的多目标优化设计、弧面蜗杆减速器的可靠性多目标优化设计、圆形振动筛减振弹簧的多目标优化设计、增压抽油泵圆柱弹簧的多目标优化设计、压缩弹簧的多目标模糊优化设计、圆柱形螺旋压缩弹簧的模糊多目标优化设计、汽车离合器拉式膜片弹簧结构参数多目标优化设计、螺旋弹簧可靠性多目标优化设计、等厚断面渐变刚度钢板弹簧的优化设计、应用模糊方法的气门弹簧可靠性多目标优化设计、圆锥滚子轴承多目标优化设计、高速仪表轴承的多目标优化设计、液体动压滑动轴承的多目标模糊优化设计、椭圆轴承多目标模糊优化设计、液体动压径向滑动轴承的概率多目标优化设计、滚动轴承的多目标优化设计、小孔节流静压轴承多目标

模糊优化设计、牙嵌式离合器多目标优化设计等研究工作。

在机械动态优化设计方面,已进行的研究工作有行星齿轮传动的动态优化设计、圆柱齿轮传动的动态分析及优化设计、舰船用齿轮传动的动态优化设计、多级齿轮传动系统动态性能的优化设计、有频率禁区的齿轮形状优化、内齿行星齿轮传动系统参数动态优化、圆弧齿双曲面齿轮降低噪声的优化设计、静压轴承自控系统动态品质的优化方法、高速柴油机气门弹簧动态优化设计等。

机械设计领域的学者还将遗传算法、神经网络等应用于机械零件的优化设计。在齿轮传动、弹簧、离合器、轴承等方面,开展了基于遗传算法的双圆弧齿轮传动的模糊可靠性优化设计、实值编码遗传算法的行星齿轮传动优化、实码遗传算法的直齿圆锥齿轮传动优化设计、门座起重机变幅机构中齿轮齿条的遗传优化、齿轮传动机构人工神经网络辅助优化设计、实数编码遗传算法的斜齿圆柱齿轮传动优化设计、径向基函数神经网络的行星齿轮减速器优化、基于遗传算法的圆锥齿轮传动模糊优化、基于遗传算法的齿轮传动离散优化设计、遗传算法的齿轮减速器优化设计、基于神经网络的变厚齿轮减速器动态优化设计、基于遗传算法的圆柱齿轮变位系数的优化选择、基于遗传算法的少齿差行星齿轮传动参数优化设计、基于遗传算法的织机变换齿轮组合优化、遗传算法的弹流润滑齿轮传动多目标优化设计、基于遗传算法的发动机气门弹簧的优化设计、基于遗传算法的弹簧的优化设计、时变回归神经网络的弹簧优化设计、汽车悬架扭杆弹簧的遗传算法优化设计、溢流阀调节弹簧遗传优化设计、基于遗传算法的普通圆柱螺旋弹簧模糊优化设计、内燃机气门压缩弹簧遗传优化设计、基于神经网络的动压轴承优化设计、基于遗传算法的液体动压滑动轴承的优化设计等研究工作。

## 2 齿轮传动优化设计

齿轮传动是机器传动系统的主要形式。以齿轮传动为主要方式的减速器等机械传动产品在工业设备领域得到了最广泛的应用。齿轮及其齿轮传动是优化技术在机械优化设计应用最早的领域。国内外学者对齿轮及其齿轮传动方面的优化问题进行了深入研究,已先后研究了齿轮传动比的优化分配、行星齿轮减速器传动比优化分配、圆柱齿轮变速器齿数比优化分配等问题,解决了齿轮参数、齿轮承载能力、齿轮弯曲强度、齿轮啮合参数、齿轮抗磨损、齿轮变位系数、齿根过渡曲线、齿轮角变位参数、齿轮修形参数等方面的优化设计问题,开展了多级平行轴圆柱齿轮减速器、斜齿轮减

速器、标准斜齿圆柱齿轮二级展开式减速器、三级齿轮传动减速器、单级圆柱齿轮减速器、双圆弧齿轮减速器、级间等强度圆锥圆柱齿轮减速器等有关齿轮传动系统的优化设计工作。

齿轮传动的优化设计包括优化模型建立和模型求解两方面问题,另外,齿轮传动还涉及例如齿轮齿数等特殊的设计变量类型问题。在齿轮传动优化设计中,有关齿轮的设计变量是离散型变量,例如整型变量的齿轮齿数、离散型变量的齿轮模数,因此在利用连续变量优化方法得到优化解后,一般还要进行必要的设计变量值的处理。逐个对各设计变量值进行圆整是常用的策略;但是这种处理方法得到结果不一定是最优解。齿轮传动的优化模型求解还可以采用混合离散变量优化方法,这样得到的解不需要进行处理,与连续变量优化方法相比,混合离散变量优化方法比较少。

齿轮强度计算是齿轮传动优化设计不可避免的问题。齿轮强度计算方法种类很多,另外各行业都有合适合行业工作特点的常规计算方法。在进行齿轮传动优化设计时,可以采用这些常规的强度计算算法,也可以采用国际标准化组织制定的标准齿轮强度计算方法。齿轮强度计算是影响齿轮工作寿命的重要因素,因此需要考虑齿轮强度计算的可靠性问题。关于齿轮强度计算的可靠性问题,可以通过对现有传动装置与新设计传动装置进行计算对比,如果计算值处于同一数量水平,则可以认为优化设计所采用的齿轮强度计算方法是可靠的。另外齿轮强度的涉及系数多,例如齿轮工作情况系数、载荷分布系数、齿宽系数、齿型系数和动载系数等,因此齿轮强度计算需要利用各种图表,有时还需要在估算基础上进行迭代才能获得有关参数值。齿轮传动优化设计中,需要对这些系数图表进行程序化。

齿轮不仅种类繁多,而且其传动的形式也复杂。本节选择最简单和常用的单级直齿圆柱齿轮减速传动为对象讨论齿轮传动的优化设计。

假设单级圆柱齿轮减速器的输入功率  $P = 280 \text{ kW}$ , 输入转速  $n_1 = 1000 \text{ r/min}$ , 传动比  $i = 5$ , 要求进行重量最轻的优化设计。

先考虑建立齿轮传动优化设计模型。按齿轮减速传动设计要求,建立重量最轻的优化目标函数。当传动齿轮和轴参数确定之后,箱体的尺寸也随即可由经验公式确定。因此有关齿轮减速传动箱的权重的主要设计变量齿轮和轴的设计参数,可按齿轮和轴的体积和最小作为优化目标。假设齿轮主传动轴长度比支承跨距  $l$  长  $28 \text{ mm}$ , 比从动轴长  $32 \text{ mm}$ , 且辐板孔数为 4, 则齿轮与轴的体积和与齿轮参数的关系如下:

$$V = \frac{\pi}{4} (d_1^3 - d_1^2) B + \frac{\pi}{4} (d_2^3 - d_2^2) B$$



$$\begin{aligned}
 & -\frac{\pi}{4}(D_2^2 - d_2^2)(B - C) - \pi d_2^2 C \\
 & + \frac{\pi}{4}(d_1^2 + d_2^2)l + \frac{\pi}{7}d_1^2 + \frac{\pi}{8}d_2^2
 \end{aligned}
 \quad (44.3-1)$$

式中

$$\begin{aligned}
 d_1 &= mz_1 & d_2 &= mz_2 i \\
 \delta &= 5m & D_2' &= mz_2 i - 10m \\
 d_{c2} &= 1.6d_2' & d_0 &= 0.25(mz_2 i - 10m - 1.6d_2') \\
 C &= 0.2B
 \end{aligned}$$

显然式(44.3-1)可写为

$$\begin{aligned}
 V &= \frac{\pi}{4}[m^2 z_1^2 B - d_1^2 B + m^2 z_2^2 i^2 B - d_2^2 B \\
 & - 0.8B(mz_2 i - 10m)^2 - 2.05Bd_2^2 \\
 & - 0.05B(mz_2 i - 10m - 1.6d_2')^2 \\
 & + d_1^2 l + d_2^2 l + 28d_1^2 + 32d_2^2] \quad (44.3-2)
 \end{aligned}$$

由上式可知,体积 $V$ 取决于 $B$ 、 $z_1$ 、 $m$ 、 $l$ 、 $d_1'$ 、 $d_2'$ 和 $i$ 七个参数,其中传动比 $i$ 为设计常量,其余六个为设计变量,因此选择六个设计变量为优化设计变量。

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} B \\ z_1 \\ m \\ l \\ d_1' \\ d_2' \end{bmatrix} \quad (44.3-3)$$

将有关参数代入式(44.3-2),经展开整理得到

$$\begin{aligned}
 f(\mathbf{x}) &= 0.78539815(4.75x_1 x_2^2 x_3^2 + 85x_1 x_2 x_4^2 \\
 & - 85x_1 x_5^2 + 0.92x_1 x_6^2 - x_1 x_5^2 \\
 & + 0.8x_1 x_2 x_3 x_6 - 1.6x_1 x_3 x_6 + x_4 x_5^2 \\
 & + x_4 x_6^2 + 28x_5^2 + 32x_6^2) \quad (44.3-4)
 \end{aligned}$$

下面考虑齿轮传动优化模型的约束条件。

齿轮传动设计要求包括强度条件、刚度条件、结构工艺条件以及尺寸和参数许用范围所规定的一些边界条件等。

(1) 有关齿轮齿数设计要求

所选择的小齿轮齿数 $z_1$ 应大于不发生根切的最小齿数 $z_{\min}$ ,对于 $h_a^* = 1.0$ , $\alpha = 20^\circ$ 的标准齿轮 $z_{\min} = 17$ ,因此这方面约束条件为

$$g_1(\mathbf{x}) = 17 - x_2 \leq 0 \quad (44.3-5)$$

(2) 有关齿轮齿宽系数设计要求

齿轮的齿宽系数荐用范围为 $16 \leq B/m \leq 35$ ,因此这方面约束条件为

$$g_2(\mathbf{x}) = 16 - \frac{x_1}{x_3} \leq 0 \quad (44.3-6)$$

$$g_3(\mathbf{x}) = \frac{x_1}{x_3} - 35 \leq 0 \quad (44.3-7)$$

(3) 有关齿轮模数设计要求

传递动力的齿轮模数一般大于2.0mm,即

$$g_4(\mathbf{x}) = 0.2 - x_3 \leq 0 \quad (44.3-8)$$

(4) 有关齿轮工艺设计要求

大齿轮的直径根据工艺条件的限制不得超过1500mm,于是小齿轮的相应直径不得超过300mm,即

$$g_5(\mathbf{x}) = x_2 x_3 - 30 \leq 0 \quad (44.3-9)$$

主、从动轴的直径范围按设计经验取: $150\text{mm} \geq d_1' \geq 100\text{mm}$ , $200\text{mm} \geq d_2' \geq 130\text{mm}$ ,则

$$g_6(\mathbf{x}) = 10 - x_5 \leq 0 \quad (44.3-10)$$

$$g_7(\mathbf{x}) = x_5 - 15 \leq 0 \quad (44.3-11)$$

$$g_8(\mathbf{x}) = 13 - x_6 \leq 0 \quad (44.3-12)$$

$$g_9(\mathbf{x}) = x_6 - 20 \leq 0 \quad (44.3-13)$$

(5) 有关轴距离设计要求

轴的支承跨距按结构关系 $l \geq B + 2\Delta_{\min} + 0.5d_1'$ ,若取 $\Delta_{\min} = 20\text{mm}$ ,则有

$$g_{10}(\mathbf{x}) = x_1 - x_4 + 0.5x_6 + 40 \leq 0 \quad (44.3-14)$$

(6) 有关齿轮强度设计要求

按齿轮的接触强度条件,有

$$\delta_c = \frac{1070}{A} \sqrt{\frac{(i+1)^3 K M n_1}{B i}} \leq [\delta_c] \quad (44.3-15)$$

式中  $K$ ——载荷系数,取 $K = 1.3$ ;

$M n_1$ ——小齿轮扭矩,按功率和转速计算得 $M n_1 = 27300\text{kg} \cdot \text{cm}$ ;

$[\delta_c]$ ——许用接触应力,按所选材料及热处理硬度取 $[\delta_c] = 8550\text{kg}/\text{cm}^2$ ;

$A$ ——齿轮啮合中心距, $A = 0.5mz_1(1+i)$ 。

将以上各量代入,经整理的约束条件为

$$g_{11}(\mathbf{x}) = \frac{441630}{x_2 x_3 \sqrt{x_1}} - 8550 \leq 0 \quad (44.3-16)$$

按小齿轮的抗弯强度条件,有

$$\delta_{w1} = \frac{2K M n_1}{B d_1 m y_1} \leq [\delta_{w1}] \quad (44.3-17)$$

式中  $d_1$ ——小齿轮分度圆直径, $d_1 = mz_1$ ;

$y_1$ ——小齿轮的齿形系数,若表示成齿数 $z_1$ 的关系式可写成

$$y_1 = 0.169 + 0.006666z_1 - 0.000854z_1^2; \quad (44.3-18)$$

$[\delta_{w1}]$ ——小齿轮许用弯曲应力,取 $[\delta_{w1}] = 261.7\text{MPa}$ 。

代入得约束条件

$$g_{12}(\mathbf{x}) = \frac{70980}{x_1 x_2 x_3^2 (0.169 + 0.006666x_2 - 0.000854x_2^2)} - 261.7 \leq 0 \quad (44.3-19)$$

按大齿轮抗弯强度条件,相应为

$$g_{13}(x) = \frac{70980}{x_1 x_2 x_3^2 (0.2824 + 0.00035395x_2 - 0.0000015765x_2^2)} - 2133 \leq 0 \quad (44.3-20)$$

(7) 有关轴强度刚度设计要求  
按主动轴的刚度条件,有

$$\frac{pl^3}{48EJ} \leq 0.003l \quad (44.3-21)$$

式中  $p$  — 齿轮齿面上作用的法向压力,  $p = 2Mn_1 / mz_1 \cos\alpha$ ;

$J$  — 惯性矩, 对轴的圆形截面  $J = \pi d_1^4 / 64$ 。

代入有

$$g_{14}(x) = \frac{0.01233x_4^3}{x_2 x_3 x_1^4} - 0.003x_1 \leq 0 \quad (44.3-22)$$

按主动轴的抗弯强度

$$\delta_{w1} = \frac{\sqrt{Mw_1^2 + (aMn_1)^2}}{W_1} \leq [\delta_{w1}] \quad (44.3-23)$$

式中  $Mw_1 = \frac{pl}{2} = \frac{2Mn_1}{mz_1 \cos\alpha} \cdot \frac{l}{2} = \frac{29200l}{mz_1}$

$$29200 \frac{x_4}{x_2 x_3}$$

$$a = [\delta_w]_r / [\delta_w]_l = 0.58$$

$$[\delta_{w1}] = [\delta_{w1}]_l = 550 \text{kg/cm}^2, W_1 = 0.1d_1^3$$

代入有

$$g_{15}(x) = \frac{1}{x_3^3} \sqrt{\left(29200 \frac{x_4}{x_2 x_3}\right)^2 + (0.58 \times 27300)^2} - 55 \leq 0 \quad (44.3-24)$$

同理, 按从动轴的抗弯强度条件有

$$g_{16}(x) = \frac{1}{x_3^3} \sqrt{\left(29200 \frac{x_4}{x_2 x_3}\right)^2 + (0.58 \times 5 \times 27300)^2} - 55 \leq 0 \quad (44.3-25)$$

因此齿轮传动优化设计数学模型为

Min.

$$\left. \begin{aligned} V &= \frac{\pi}{4} [m^2 z_1^2 B - d_1^2 B + m^2 z_2^2 i^2 B \\ &\quad - d_2^2 B - 0.8B(mz_1 - 10m)^2 + 2.05Bd_2^2 \\ &\quad - 0.05B(mz_1 i - 10m - 1.6d_2^2) + d_1^2 l \\ &\quad + d_2^2 l + 28d_1^2 + 32d_2^2] \\ \text{s.t.} \quad &g_u(x) \leq 0 \quad (u=1, 2, 3, \dots, 16) \end{aligned} \right\} \quad (44.3-26)$$

采用罚函数法对齿轮传动优化设计模型进行求

解。取原设计的减速器参数方案作初始点, 即

$$x^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ x_4^{(0)} \\ x_5^{(0)} \\ x_6^{(0)} \end{bmatrix} = \begin{bmatrix} B^{(0)} \\ z^{(0)} \\ m^{(0)} \\ l^{(0)} \\ d_1^{(0)} \\ d_2^{(0)} \end{bmatrix} = \begin{bmatrix} 23 \\ 21 \\ 0.8 \\ 42 \\ 12 \\ 16 \end{bmatrix}$$

取初始罚因子为 5596, 惩罚因子的降低系数取  $C = 0.5$ 。在共轭方向法无约束优化过程中的一维搜索用二次插值法, 精度取  $\epsilon = 10^{-7}$ 。共轭方向无约束优化和最后的收敛精度取  $\delta = 10^{-3}$ 。经过 38 次迭代后得到优化解, 优化解为

$$x^* = \begin{bmatrix} B^* \\ z^* \\ m^* \\ l^* \\ d_1^* \\ d_2^* \end{bmatrix} = \begin{bmatrix} 13.0929 \\ 18.7388 \\ 0.8183 \\ 23.5930 \\ 10.0001 \\ 13.0000 \end{bmatrix}$$

$$f(x^*) = 35334.3585 \text{cm}^3$$

齿轮传动优化计算结果还有待修正, 例如齿轮  $z$  应该取为整数, 模数应取为标准值。

### 3 弹簧优化设计

在进行弹簧设计时, 一般需要根据最大工作载荷、最大变形等设计要求来进行弹簧丝直径、弹簧中径、弹簧工作圈数、弹簧螺旋升角、弹簧高度等参数的计算。弹簧设计有时需要多次迭代才能得到满足设计要求的参数, 因此有必要采取优化技术进行弹簧的设计。目前国内已经开展了压缩弹簧、碟形弹簧、波形弹簧、板弹簧、片弹簧等不同类型弹簧的优化设计工作。弹簧零件结构类型、产品种类繁多, 本节介绍弹簧中最常用的普通圆柱压缩螺旋弹簧的优化设计。

进行普通圆柱压缩螺旋弹簧设计, 是在给定的最大工作载荷和所要求的变形量情况下选择材料、确定出许用剪应力, 计算弹簧的弹簧丝的直径、弹簧中径、弹簧的有效工作圈数、材料的切变弹性模量、旋绕比、曲度系数等结构参数。在弹簧设计计算过程中应满足强度和变形方面的要求。

在弹簧设计时, 为了保证弹簧的正常工作和有足够的寿命, 还应对弹簧的疲劳强度、稳定性等方面指标进行验算。如果弹簧工作在变剪应力环境时, 可根据弹簧受到的载荷等进行弹簧的疲劳强度验算。根据弹簧自由高度、弹簧的中径、弹簧的刚度、长度折算系数等, 可进行弹簧的稳定性验算。

弹簧的优化设计, 一般选择弹簧的弹簧丝直径、中径、工作圈数等结构参数作为优化设计变量, 用弹簧的

承载能力、变形、刚度、抗拉强度等性能参数作为约束条件,弹簧的重量、体积、寿命等指标可作为优化的目标。

假设弹簧的初始设计参数为:安装高度  $H_1 = 50.8\text{mm}$ , 初始载荷  $F_1 = 272\text{N}$ , 最大工作载荷  $F_2 = 680\text{N}$ , 工作行程  $h = 10.16\text{mm}$ , 弹簧的工作频率  $f_c = 50\text{Hz}$ ; 弹簧丝用油淬回火的 50CrVA 钢丝, 进行喷丸处理; 工作温度为  $126^\circ\text{C}$ ; 要求弹簧中径为  $20\text{mm} \leq D_2 \leq 50\text{mm}$ , 弹簧总圈数为  $4 \leq n_1 \leq 50$ , 支承圈数  $n_2 = 1.75$ , 旋绕比  $C \geq 6$ ; 安全系数为 1.2。要求进行弹簧重量最轻优化设计。

先建立弹簧的优化设计模型。选择弹簧钢丝直径  $d$ 、弹簧中径  $D_2$  和弹簧总圈数  $n_1$  为优化设计变量, 即

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d \\ D_2 \\ n_1 \end{bmatrix} \quad (44.3-27)$$

选择弹簧的重量为优化设计的目标函数。弹簧的重量函数为

$$W = \frac{\pi^2}{4} d^2 D_2 n_1 \rho \quad (44.3-28)$$

式中  $\rho$ — 钢丝材料的密度,  $\rho = 7.8 \times 10^{-6} \text{kg/mm}^3$  因此弹簧的优化目标函数为

$$\begin{aligned} f(\mathbf{x}) &= W = \frac{\pi^2}{4} \times 7.8 \times 10^{-6} x_1^2 x_2 x_3 \\ &= 0.1925 \times 10^{-3} x_1^2 x_2 x_3 \end{aligned}$$

依据对弹簧功能和结构的设计要求, 可建立如下弹簧优化设计模型的约束条件:

(1) 疲劳强度方面的设计要求

$$g_1(\mathbf{x}) = \frac{S_{\min} - S}{S} = \frac{1.2}{S} - 1 \leq 0 \quad (44.3-29)$$

式中取  $S_{\min} = 1.2$ 。

因  $\tau_a = 0.5\sigma_b = 740\text{N/mm}$ ,  $\tau_0 = 365.4\text{N/mm}$  得

$$\begin{aligned} S &= \frac{\tau_0}{\left(\frac{2\tau_a - \tau_0}{\tau_1}\right) \tau_a + \left(\frac{\tau_1}{\tau_0}\right) \tau_m} \\ &= \frac{365.4}{1.513\tau_a + 0.487\tau_m} \end{aligned}$$

式中得剪应力幅  $\tau_a$  为

$$\begin{aligned} \tau_a &= k \frac{8F_a D_a}{\pi d^3} = \frac{16}{(D_2/d)^3} \cdot \frac{8F_a D_2}{\pi d^3} \\ &= \frac{830.3x_2^{0.46}}{x_1^{2.86}} \end{aligned}$$

平均应力幅  $\tau_m$  为

$$\begin{aligned} \tau_m &= \left(1 + \frac{0.615d}{D_2}\right) \frac{8F_m D_2}{\pi d^3} \\ &= \frac{1212.12x_2}{x_1^3} + \frac{745.46}{x_1^2} \end{aligned}$$

(2) 弹簧旋绕比方面的设计要求

$$g_2(\mathbf{x}) = \frac{C_{\min} - C}{C} = \frac{6x_1}{x_2} - 1 \leq 0$$

(3) 弹簧中径尺寸方面的设计要求

$$g_3(\mathbf{x}) = \frac{D_{\min} - D_2}{D_2} = \frac{20}{x_2} - 1 \leq 0 \quad (44.3-30)$$

$$g_4(\mathbf{x}) = \frac{D_2 - D_{\max}}{D_{\max}} = \frac{x_2}{50} - 1 \leq 0 \quad (44.3-31)$$

(4) 弹簧的稳定性方面设计要求

$$g_5(\mathbf{x}) = \frac{F_2 - F_1}{F_c} = \frac{680}{F_c} - 1 \leq 0 \quad (44.3-32)$$

式中  $F_c$  (按一端固定、一端铰支考虑,  $\mu$  为 0.7) 为

$$\begin{aligned} F_c &= 32.68[(x_3 - 0.5)x_1 + 20.304] \times \\ &\left\{1 - \sqrt{1 - 13.980 \left[\frac{x_2}{(x_3 - 0.5)x_1 + 20.304}\right]^2}\right\} \end{aligned}$$

(5) 弹簧共振性设计条件

$$g_6(\mathbf{x}) = \frac{13f_c - f}{f} = \frac{650}{f} - 1 \leq 0 \quad (44.3-33)$$

式中, 弹簧自振频率为

$$f = 3.56 \times 10^5 \frac{d}{n_1 D_2^2} = 3.56 \times 10^5 \frac{x_1}{x_3 x_2^2}$$

(6) 弹簧其他设计要求

为了保证弹簧在最大载荷作用下不发生并圈现象, 要求弹簧在最大载荷  $F_2$  时得高度  $H_2$  应大于压并高度  $H_b$ 。为了保证弹簧具有足够的刚度, 要求弹簧的刚度  $K_a$  与设计要求的刚度  $K$  的误差小于 1/100。

这方面约束条件为

$$g_7(\mathbf{x}) = \frac{H_b - H_2}{H_2} = 0.246x_1 x_3 - 0.0123x_1 - 1 \leq 0 \quad (44.3-34)$$

$$h(\mathbf{x}) = \begin{cases} 0 & \text{当 } (\theta) \leq 0 \\ (\theta) & \text{当 } (\theta) > 0 \end{cases} \quad (44.3-35)$$

式中  $(\theta)$ — 误差值, 即

$$(\theta) = |k_a - k_i - k/100|$$

$$= \left| \frac{Gx_1}{8x_2^3(x_1 - 1.75)} \quad 4.02 \right| \quad 0.0402$$

因此以重量为目标的弹簧的优化设计模型为

Min.

$$\left. \begin{aligned} f(x) &= W = \frac{\pi^2}{4} \times 7.8 \times 10^{-6} x_1^2 x_2 x_3 \\ &= 0.1925 \times 10^{-4} x_1^2 x_2 x_3 \\ x &= [x_1, x_2, x_3]^T \\ \text{s. t.} \\ g_u(x) &\leq 0 \quad (u = 1, 2, \dots, 7) \\ h(x) &= \begin{cases} 0 & \text{当 } (\theta) \leq 0 \\ (\theta) & \text{当 } (\theta) > 0 \end{cases} \end{aligned} \right\} \quad (44.3-36)$$

采用求解约束优化问题直接解法——随机方向搜索法计算弹簧优化设计模型。优化模型的优化参数为， $20\text{mm} \leq x_2 \leq 50\text{mm}$  和  $4 \leq x_3 \leq 50$ ，初始点  $x^{(0)} = [6.40, 6.7873]^T$ ，初始步长  $\alpha^{(0)} = 0.05$ 。优化计算得到的优化解如表 44.3-1 所示。

表 44.3-1 弹簧优化设计解

钢丝直径 /mm	弹簧中径 /mm	总圈数	自由高度 /mm	弹簧重量 /kg
5.8295	37.1729	7.3370	57.5615	0.1784

### 4 轴承优化设计

轴承是支撑旋转零件的装置，被广泛应用于各种机械设备中。轴承种类繁多，按轴承承载载荷方向的不同，轴承可分为径向轴承和推力轴承；按轴承工作时摩擦性质不同，轴承可分为滑动轴承和滚动轴承。所谓径向轴承，就是轴承上的反作用力与轴心线垂直的轴承。所谓推力轴承，就是轴承上的反作用力与轴心线方向一致的轴承。轴承设计中包含了一些优化问题，如滚动轴承额定寿命的优化选型、球面轴承结构参数的优化、液体动压润滑轴承的优化设计、传动轴承结构参数的优化设计、螺旋油楔动压轴承的优化设计、空气静压轴承的优化设计、调心滚子轴承的优化设计等。本节主要讨论液体动压滑动轴承的优化设计问题。

液体动压滑动轴承的参数包括几何形状参数、润滑油的性能参数、载荷特性参数等。为了确保液体动压滑动轴承处于稳定的液体润滑工作状态，轴承在设计时还应考虑满足油膜厚度的限制、温升的限制、润滑油最小粘度、最大压力等方面设计条件。

液体动压滑动轴承的优化设计同样包括优化设计数学模型建立和优化模型求解二方面内容。液体动压滑动轴承的优化设计指标可以是最低的温升、最少润滑油用量、最小摩擦损失等，也可以是这些指标的组

假设进行最少润滑油用量和升温最小的轴承优化设计。液体动压滑动轴承的设计参数如下：

轴承所受的载荷分别为： $F = 8896\text{N}$ 、 $4448\text{N}$ 、 $2224\text{N}$ 、 $1112\text{N}$ 。工作转速分别为  $n = 1000\text{r/min}$  ( $16.66\text{r/s}$ )、 $2000\text{r/min}$  ( $33.33\text{r/s}$ )、 $5000\text{r/min}$  ( $83.33\text{r/s}$ )、 $10000\text{r/min}$  ( $166.66\text{r/s}$ )、 $15000\text{r/min}$  ( $250\text{r/s}$ )、 $20000\text{r/min}$  ( $333.33\text{r/s}$ )，轴承的直径  $D = 50.8\text{mm}$ 。

轴承的设计条件为

$$\left. \begin{aligned} [h_{\min}] &= 1.27 \times 10^{-3} \text{mm}, [t_{\max}] = 50^\circ\text{C} \\ [P] &\leq 207\text{N/mm}^2 \\ 0.25 &\leq L/D \leq 1 \\ \mu_{\min} &= 6.895 \times 10^{-4} \text{Pa} \cdot \text{s} \end{aligned} \right\}$$

先建立轴承的优化设计模型。选择轴承的径向间隙  $C$ 、宽径比  $L/D$  和润滑油的粘度  $\mu$  为设计变量，即

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} C \\ L/D \\ \mu \end{bmatrix} \quad (44.3-37)$$

这样可以将有关轴承油膜厚度、温升限制、润滑油最小粘度、最大压力、轴承宽度、轴承运转稳定性方面的设计基本公式和参数表示为如下函数式：

$$S = \left( \frac{R}{C} \right)^2 \frac{\mu n}{P} = \left( \frac{R}{x_1} \right)^2 \frac{n x_3}{P} \quad (44.3-38)$$

$$h_{\min} = \varphi_1(x) \quad (44.3-39)$$

$$\Delta t = \varphi_2(x) \quad (44.3-40)$$

$$\frac{P}{P_{\max}} = \varphi_3(x) \quad (44.3-41)$$

$$\frac{Q}{RnDx_1x_2} = \varphi_4(x) \quad (44.3-42)$$

$$f \frac{R}{x_1} = \varphi_5(x) \quad (44.3-43)$$

$$\frac{MC\omega^2}{F} = \frac{M\omega^2x_1}{F} \leq \psi(\sigma) = \pi S x_2^2 \quad (44.3-44)$$

按轴承最少用油量和最小的温升来建立目标函数，并将二项指标按照按 5:1 加权因子进行加权，由此可得目标函数

$$f(x) = 5Q + \Delta t = 5RnDx_1x_2\varphi_4(x) + \varphi_2(x) \quad (44.3-45)$$

下面根据轴承结构、性能和润滑油粘度等方面设计要求建立轴承优化设计模型的约束条件。

(1) 最小油膜厚度设计要求

为了保证轴承形成液体摩擦润滑，应使轴承的最小油膜厚度大于允许的最小油膜厚度，即

$$h_{\min} \geq [h_{\min}] \quad \text{或} \quad \varphi_1(x) \geq 1.27 \times 10^{-3} \text{mm} \quad (44.3-46)$$

从而得约束条件

$$g_1(x) = \frac{1.27 \times 10^{-3} - \varphi_1(x)}{\varphi_1(x)} \leq 0 \quad (44.3-47)$$

## (2) 升温方面设计要求

根据设计要求,应尽量减少轴承升温,在不得已的情况下也不能超过给定的升温。轴承的最高允许温度为 50°C, 设进油温度为 35°C, 则轴承的允许温升  $[\Delta t] = 15^\circ\text{C}$ 。因此, 轴承的升温应满足

$$\Delta t \leq [\Delta t] \quad \text{或} \quad \varphi_2(x) \leq 15^\circ\text{C} \quad (44.3-48)$$

从而可得升温约束条件

$$g_2(x) = \frac{\varphi_2(x) - 15}{15} \leq 0 \quad (44.3-49)$$

## (3) 轴承比压方面设计要求

为了保证油膜不被破坏, 轴承的最大比压应小于允许值, 即

$$P_{\max} \leq [P] \quad \text{或} \quad \frac{P}{\varphi_3(x)} \leq 207\text{N/mm}^2 \quad (44.3-50)$$

从而得到比压约束条件

$$g_3(x) = \frac{[P/\varphi_3(x)] - 20.7}{20.7} \leq 0 \quad (44.3-51)$$

## (4) 润滑油粘度方面设计要求

根据设计要求, 轴承所使用的润滑油粘度应大于允许的最小粘度, 即

$$\mu \geq [\mu_{\min}] \quad \text{或} \quad x_3 \geq 6.895 \times 10^{-4}\text{Pa} \cdot \text{s} \quad (44.3-52)$$

从而得到如下约束条件,

$$g_4(x) = \frac{6.895 \times 10^{-4} - x_3}{x_3} \leq 0 \quad (44.3-53)$$

## (5) 轴承宽径比方面设计要求

根据设计要求, 轴承的宽径比应满足  $0.25 \leq L/D \leq 1$  的条件, 由此得到约束条件

$$g_5(x) = \frac{0.25 - x_2}{x_2} \leq 0 \quad (44.3-54)$$

$$g_6(x) = \frac{x_2 - 1.0}{1.0} \leq 0 \quad (44.3-55)$$

## (6) 轴承运转稳定性方面设计要求

为了保证轴承运转的稳定性, 应满足如下条件

$$\frac{M\omega^2 x_1}{F} \leq \pi S x_2^2 \quad (44.3-56)$$

或 
$$\frac{M\omega^2 x_1}{F} \leq \left(\frac{R}{x_1}\right)^2 \frac{n\pi x_2^2 x_3}{P}$$

从而得到运转稳定性的约束条件

$$g_7(x) = \frac{(M\omega^2 x_1/F) - [(R/x_1)^2 n\pi x_2^2 x_3/P]}{(R/x_1)^2 n\pi x_2^2 x_3/P} \leq 0 \quad (44.3-57)$$

因此可以建立包含 3 个设计变量、7 个约束条件液体动压滑动轴承的优化设计模型。在液体动压滑动轴承的优化模型中, 约束条件  $g_1(x) \sim g_6(x)$  主要是限制变量的变化范围, 属于设计变量边界约束条件。

液体动压滑动轴承的优化设计模型如下式所示:

$$\left. \begin{aligned} & \text{Min.} \\ & f(x) = 5Q + \Delta t = 5RnDx_1 x_2 \varphi_4(x) + \varphi_2(x) \\ & x = [x_1, x_2, x_3]^T \\ & \text{s.t.} \quad g_i(x) \leq 0 \quad (i=1, 2, \dots, 7) \end{aligned} \right\} \quad (44.3-58)$$

分别四种工作载荷和六种工作转速情况进行优化计算, 得到了不同工作状况下的设计变量轴承的径向间隙、宽径比和润滑油粘度的优化解。

## 第4章 机械系统优化设计

### 1 机械系统优化设计概述

所谓机械系统就是指为了实现特定功能的、相互间具有联系的若干个零件、部件和装置构成的整体。现代机械系统功能多,构成也十分复杂。机械系统一般包括动力、传动、执行、操纵及控制等部分。现代机械系统除满足功能、强度、刚度、寿命、体积、重量等方面要求外,还需要考虑机械系统的可靠性、安全性、经济性,以及噪声和振动等有关环境等方面的许多问题。因此机械系统的设计工作越来越复杂,同时也越来越重要。一般可以认为机械系统设计质量和经济等方面指标取决于设计的质量,没有高质量的设计,就不可能有高质量的机械系统。如果在机械系统设计阶段,没有考虑到具体的功能要求,那么无论在经过怎样的制造、安装和调试,都无法弥补机械系统设计时的缺陷。因为机械系统的功能等要求是由设计决定的,制造、管理等其他生产环节只能对机械系统产品质量起到保证作用。

在机械系统设计中,机械系统总体设计是重要内容之一,通过机械系统的总体设计,可以确定影响机械性能的总体参数。目前机械系统设计与传统的设计有着明显的差别。传统的设计重点在于零件或产品内部的设计问题,主要考虑零件或产品的内部性能等;机械系统设计重点在于对组成系统的各零部件之间、机械系统与周围运行环境等外部系统之间的设计问题,主要考虑零件之间、机械内部系统和外部系统之间相互作用和影响的性能等。机械系统设计采用系统的设计观点,通过组成系统的各零部件之间协调来实现整个机械系统最佳的性能设计,因此机械系统设计过程中有必要采取优化设计的思想。一般机械系统也可以看作是由确定的质量、刚度和阻尼的物体组成并能完成特定功能的系统,因此动态性能是机械系统的重要特性。针对机械系统动态性能问题,一些学者提出了机械系统动态优化设计。近若干年来,在机械系统领域开展了一些优化方法和优化设计的研究工作。例如,大型机械系统整体优化策略的研究、复杂机械系统优化设计的一种策略、基于 Hopfield 神经网络的机械系统优化研究、机械系统可靠度的多目标模糊优化决策、机械系统可靠性的多目标模糊优化设计、配气机构机械系统的多目标优化及模糊解法、机械系统整体可靠性优化设计的分解协调法、机械系统可靠性优化设计的实现

策略、串联机械系统可靠性优化设计的目标协调法、非平稳随机激励的机械系统动态优化设计方法、机械系统动态优化设计的混合遗传基因方法、机械系统动态优化设计的复合遗传算法、机械系统结构优化设计的遗传基因法、机械系统动力学参数优化设计的遗传算法、遗传算法在机械系统动力学优化设计的应用等。

本章先介绍机械系统动态优化设计方法,机电系统是机械系统设计时经常遇到的问题,因此本章也包括了机电系统的优化设计方法。

### 2 机械系统动态优化设计

机械系统动态优化设计是指将优化方法与机械系统动态性能数值计算方法相结合面形成的设计方法。机械系统动态优化设计是随着静态结构系统分析、优化理论的发展而逐步形成的,目前机械系统动态优化设计研究对象从梁、板、壳发展到了柔体结构,从简单系统发展到了复杂系统。利用机械动态优化设计方法,可使得机械系统在设定的限制条件下得到最佳的动态性能。与传统的设计方法相比,机械系统动态优化设计具有明显的优点。传统设计在满足动态性能指标方面通过动力性能实验、修改系统参数、再实验等方式实现的,其设计方法周期长、费用高、效率低,有时机械系统动态性能还不能满足。

机械动态优化设计是以动态分析和动态设计为基础的。目前大多数商业有限元分析软件都支持机械系统的动态分析功能;利用有限元分析软件,可进行结构或系统的包括固有频率和振型参数在内的模态分析,可进行结构或系统的包括速度、加速度、位移与振幅在内的系统响应分析,可进行结构或系统的时变或非时变激励分析等。系统动态分析一般以系统输入、系统和系统输出为研究对象,进行系统输入、系统响应和系统数学模型分析。以系统输入和系统特性为基础,利用动态分析可以进行系统动态输出响应计算;以系统输出和系统特性为基础,可以进行系统的输入分析;以系统输入和系统输出为基础,可以进行系统的固有特性分析。动态设计是以动态分析为基础计算系统的设计参数。在动态设计过程中,如果系统的动态性能得不到满足,则修改系统设计参数。机械系统动态优化设计可以认为是以动态设计和优化算法为基础的设计方法。采用优化算法进行动态优化设计,可以利用动态分析来计算机械系统中与动态性能相关的目标函数、约束函

数和目标及约束函数的梯度信息，并进行系统设计参数的优选。

机械系统的动态特性可用公式 (44.4-1) 来描述：

$$\left. \begin{aligned} [M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} &= \{P(t)\} \\ \{x(0)\} &= \{x_0\} \\ \{\dot{x}(0)\} &= \{\dot{x}_0\} \end{aligned} \right\} \quad (44.4-1)$$

式中  $[M]$ 、 $[C]$ 、 $[K]$ ——分别是系统质量、阻尼和刚度矩阵；

$\{P(t)\}$  激振力向量。

机械系统动态优化设计数学模型可以描述为，确定系统参数  $[M]$ 、 $[C]$ 、 $[K]$ ，使目标函数  $F = f(t, x(t))$ ， $[M]$ 、 $[C]$ 、 $[K]$  在满足时间区间  $[0, T]$  内动态性能约束条件

$$g_i(t, x(t), [M], [C], [K]) \leq 0 \quad i=1, 2, \dots, m \text{ 时} \\ \text{最小, 即} \\ \text{Min.}$$

$$F = f(t, x(t), [M], [C], [K]) \\ X = \{[M], [C], [K]\} \quad (44.4-2)$$

s. t.

$$g_i(t, x(t), [M], [C], [K]) \leq 0 \quad i=1, 2, \dots, m$$

在式 (44.4-2) 描述的动态优化模型中，目标函数和约束条件都与系统动态性能参数有关。因为优化迭代过程需要求解微分方程描述的动态性能，因此系统动态性能的计算是机械系统动态优化的关键问题；有时对于包括强非线性因素的机械动态系统，可能会由于系统动态分析问题而难以得到优化解。

动态优化设计问题可分为稳态约束条件优化设计问题、瞬态约束条件优化设计问题、时变约束优化设计问题和非时变约束条件优化设计问题。稳态约束条件优化设计方法相对成熟，瞬态响应条件优化设计问题求解相对困难。

下面以齿轮传动系统动态性能优化设计为例介绍动态性能优化设计。

齿轮传动的动态性能优化设计是有关一个啮合周期内的动态性能指标的优化设计。在进行齿轮动态优化设计前需要建立齿轮振动模型，并对振动模型进行求解得到动态响应，然后建立动态优化数学模型，最后进行优化计算。

齿轮传动系统的动态性能指标包括最大动载荷、动载系数、齿轮刚度值、最大振动加速度等几种。

所谓最大动载荷，就是指齿轮啮合时在一个啮合周期内由振动在轮齿间作用而产生的最大载荷。最大

动载荷可由下式计算：

$$F_{dmax} = \max_{i=1, \dots, n} [k_i(x_i - e_i)] \quad (44.4-3)$$

式中  $k_i, x_i$ ——分别为啮合位置  $i$  处的单对齿刚度、相对振动位移；

$e_i$  ——在啮合位置  $i$  处的齿轮误差；

$n$  ——一个啮合周期内的等分点数。

所谓动载系数，就是一个啮合周期内轮齿所受的最大动载荷与法向静载荷之比，即

$$K_v = F_{dmax}/F_n = \max_{i=1, \dots, n} [k_i(x_i - e_i)]/F_n \quad (44.4-4)$$

齿轮刚度方面的指标包括齿轮副的平均刚度、啮合刚度变化幅度均方根值等。所谓齿轮副的平均刚度，就是齿轮在一个啮合周期内啮合线上各个位置的轮齿刚度的平均值。根据该值大致可算出齿轮的共振区域。啮合刚度变化幅度均方根值计算式为

$$\bar{K} = \sum_{i=1}^n \frac{\Delta r(K_i + K_{i+1})/2}{T_z} \quad (44.4-5)$$

齿轮刚度变化幅度均方根值反映齿轮啮合的振动冲击程度。轮齿刚度变化幅度均方根值计算公式为

$$K_r = \left[ \sum_{i=1}^n (K_i - \bar{K})^2/n \right]^{1/2} \quad (44.4-6)$$

齿轮的振动载荷基本上与振动加速度成比例，因此，振动加速度是常用的动态性能指标。由于齿轮的噪声与齿轮振动加速度均方值成正比，因此，扭转振动加速度均方根值可近似表示齿轮的振动和噪声的大小，其计算式为

$$\bar{a} = \left[ \sum_{i=1}^n \ddot{x}_i/n \right]^{1/2} \quad (44.4-7)$$

下面介绍齿轮传动系统动态优化设计的数学模型建立。

齿轮传动的设计参数有齿数  $Z$ 、模数  $m$ 、变位系数  $\chi_{n1}$  和  $\chi_{n2}$ 、齿宽  $b$ 、分度圆螺旋角  $\beta$  及齿形修整量  $e_n$  等。对于动态性能优化设计的优化设计变量，可选择对齿轮动态性能影响较大的参数作为设计变量，因此将齿数  $Z$ 、齿宽  $b$ 、分度圆螺旋角  $\beta$ 、变位系数  $\chi_{n1}$  和  $\chi_{n2}$ 、齿顶高系数  $h_a^*$ 、齿形修整量  $e_n$  等作为设计变量，其余的设计参数作为常量，优化设计变量为

$$X = [Z, b, \beta, \chi_{n1}, \chi_{n2}, h_a^*, e_n]^T$$

在齿轮传动系统动态优化设计中，优化目标函数可采用啮合周期内最大动态性能指标，例如可选择本节介绍的齿轮的最大动载荷、动载系数等其中的任一指标作为设计目标函数。齿轮传动系统动态优化设计中，约束函数主要包括：

## (1) 边界约束

$$\begin{aligned} g_1[X] &= x_1^l/x_1 - 1 \leq 0 \\ g_j[X] &= 1 - x_j^l/x_j \leq 0 \quad i=1,2,\dots,7; j=8,9,\dots,14 \end{aligned} \quad (44.4-8)$$

式中  $x_i^l$  —— 设计变量的上界;

$x_i^l$  —— 设计变量的下界。

## (2) 静态性能约束

## 1) 不发生根切的约束函数:

$$\begin{aligned} g_{15}[X] &= h_a^* - Z_1 \sin^2 \alpha_i / (2 \cos \beta) - \chi_{n1} \leq 0 \\ g_{16}[X] &= h_a^* - Z_2 \sin^2 \alpha_i / (2 \cos \beta) - \chi_{n2} \leq 0 \end{aligned} \quad (44.4-9)$$

式中  $Z_1, Z_2$  —— 分别为小齿轮和大齿轮的齿数;

$\alpha_i$  —— 分度圆端面压力角。

## 2) 齿顶厚度的限制条件:

为了保证齿顶有足够的强度,齿顶厚度应大于 0.25m,则约束函数为

$$\begin{aligned} g_{17}[X] &= 0.25m_n / \cos \beta - s_1 d_{a1} / d_1 + d_{a1} (\operatorname{inv} \alpha_{a1} \operatorname{inv} \alpha_i) \\ g_{18}[X] &= 0.25m_n / \cos \beta - s_2 d_{a2} / d_2 + d_{a2} (\operatorname{inv} \alpha_{a2} \operatorname{inv} \alpha_i) \end{aligned} \quad (44.4-10)$$

式中  $s_1, s_2$  —— 分别为小齿轮和大齿轮的分度圆齿厚;

$d_{a1}, d_{a2}$  —— 分别为小齿轮和大齿轮的顶圆直径;

$d_1, d_2$  —— 分别为小齿轮和大齿轮的分度圆直径。

## 3) 重合度的约束函数:

为了保证齿轮副的连续传动条件,重合度应满足  $1.2 \leq \epsilon_a \leq 2$ ,则约束函数为

$$\begin{aligned} g_{19}[X] &= 1.2 - [Z_1 (\tan \alpha_{a1} - \tan \alpha_i') + \\ &\quad Z_2 (\tan \alpha_{a2} - \tan \alpha_i')] / (2\pi) \leq 0 \\ g_{20}[X] &= [Z_1 (\tan \alpha_{a1} - \tan \alpha_i') + \\ &\quad Z_2 (\tan \alpha_{a2} - \tan \alpha_i')] / (2\pi) - 2 \leq 0 \end{aligned} \quad (44.4-11)$$

式中  $\alpha_{a1}, \alpha_{a2}$  —— 分别为小齿轮和大齿轮的顶圆压力角;

$\alpha_i'$  —— 端面啮合角。

## 4) 不发生过渡曲线干涉的条件:

小齿轮不发生过渡曲线干涉的约束函数为

$$\begin{aligned} g_{21}[X] &= \tan \alpha_i - \tan \alpha_i' + Z_2 (\tan \alpha_i - \tan \alpha_i') / Z_1 \\ &\quad - 4(h_a^* - \chi_{n1}) \cos \beta / (Z_1 \sin 2\alpha_i) \leq 0 \\ g_{22}[X] &= \tan \alpha_i - \tan \alpha_i' + Z_2 (\tan \alpha_i - \tan \alpha_i') / Z_2 \\ &\quad - 4(h_a^* - \chi_{n2}) \cos \beta / (Z_2 \sin 2\alpha_i) \leq 0 \end{aligned} \quad (44.4-12)$$

## 5) 齿面接触强度的约束函数:

齿面接触强度条件:

$S_{H1} \geq S_{Hmin}$  ( $i=1,2$ ),则约束函数为

$$\begin{aligned} g_{23}[X] &= S_{Hmin} / S_{H1} - 1 \leq 0 \\ g_{24}[X] &= S_{Hmin} / S_{H2} - 1 \leq 0 \end{aligned} \quad (44.4-13)$$

式中  $S_{H1}, S_{H2}$  —— 分别为小齿轮和大齿轮的接触强度计算安全系数;

$S_{Hmin}$  —— 齿面接触强度的许用最小安全系数。

## 5) 齿根弯曲强度的约束函数:

齿根弯曲强度条件:  $S_{F1} \geq S_{Fmin}$  ( $i=1,2$ ),则约束函数为

$$\begin{aligned} g_{25}[X] &= S_{Fmin} / S_{F1} - 1 \leq 0 \\ g_{26}[X] &= S_{Fmin} / S_{F2} - 1 \leq 0 \end{aligned} \quad (44.4-14)$$

式中  $S_{F1}, S_{F2}$  —— 分别为小齿轮和大齿轮的弯曲强度计算安全系数;

$S_{Fmin}$  —— 弯曲强度的许用最小安全系数。

## (3) 动态性能约束函数

在齿轮传动系统动态优化设计中,可将已作为目标函数以外的性能指标作为约束条件。

齿轮传动系统动态优化设计的数学模型为

$$\min. f[X] = \max_{i=1, \dots, n} [k_i(x_i - e_i)]$$

$$\text{或} \quad \min. f[X] = \left[ \sum_{i=1}^n \bar{x}_i / n \right]^{1/2} \quad (44.4-15)$$

.....

$$X = [Z_1, b, \beta, \chi_{n1}, \chi_{n2}, h_a^*, e_n]^T$$

$$\text{s. t.} \quad g_i[X] \leq 0 \quad i=1,2,\dots,29$$

为了降低优化难度,进一步减少了设计变量数目,仅选取变位系数  $\chi_{n1}$  和  $\chi_{n2}$ 、齿顶高系数  $h_a^*$ 、齿形修整量  $e_n$  为设计变量,而齿数、模数和齿宽用常规方法确定,取  $Z_1=Z_2=25, m=5\text{mm}, b=14\text{mm}$ 。采用约束拟牛顿法的连续变量优化方法进行求解,变位系数、齿顶高系数和齿形修整量的优化解分别为 0.151、0.078、1.192 和 0.008,优化目标函数值为 60.287。

在机械系统动态优化设计中,由于动态优化问题的复杂性应尽可能简化优化模型,减少设计参数数目。在对优化模型中有关动态性能指标计算时,需要建立有限元动态分析模型,采用高效、可靠、稳定的动态分析方法。如果利用商业化的有限元分析软件进行动态性能分析,需要注意有限元网格自动划分及有限元动态分析模型与优化算法接口问题,从而实现目标函数和约束函数的方便计算。如果优化算法是基于梯度的算法,则还必须提供目标函数、约束函数的梯度计算模块与优化算法的接口。通常的商业有限元软件不提供目标函数、约束函数的梯度计算模块。

## 3 机电系统优化设计

机电一体化是复杂机械系统的特征之一,早在 20



世纪 70 年代,日本学者就提出了机电一体化概念。所谓机电一体化就是指机械与电子技术的综合集成,是将电子、信息技术应用于传统的机械装备,以工业机器人、微电子制造设备、自动检测与测量设备为代表的复杂机械系统是典型的机电相互结合的系统。显然机电相互结合的复杂机械系统存在设计难度大等问题。在传统的机电系统设计过程中,首先对机械部分进行设计,然后再进行控制部分的设计。在机电系统中,机械部分的动力学特性对控制部分有着重要的影响,而控制部分对机械部分的运动特性也有着影响,因此机电系统机械结构参数和控制参数之间存在着很强的耦合关系。因此传统的机电系统设计方式虽然降低了系统的复杂性和设计的难度,但是很难解决机电系统明显间存在的耦合问题,尤其对高精度机电系统设计更为困难。

目前在机电系统设计方面已经开展了有些工作,例如分析固定结构与控制的机电耦合设计方法,用回归实验优化方法对机电系统进行设计的策略。这里以单柔性臂机构为例介绍了机电系统优化设计方法。

机电系统动力学的各种特征和品质指标在很大程度上是由系统的极点决定的。因此,可以根据系统的性能要求预先选择极点的位置,然后通过对系统优化来设计系统的其他参数。

图 44.4-1 为柔性梁的变形示意图及相应的坐标系,在图中把梁作为一根质量沿长度方向均匀分布的杆件,设梁的长度为  $L$ ,截面对  $Y$  轴的抗弯刚度为  $EI_y$ ,质量密度为  $\rho$ ,截面积为  $A$ ,电机扭矩为  $T$ , $C$  为扭矩  $T$  的作用点位置到回转中心  $O$  的距离。

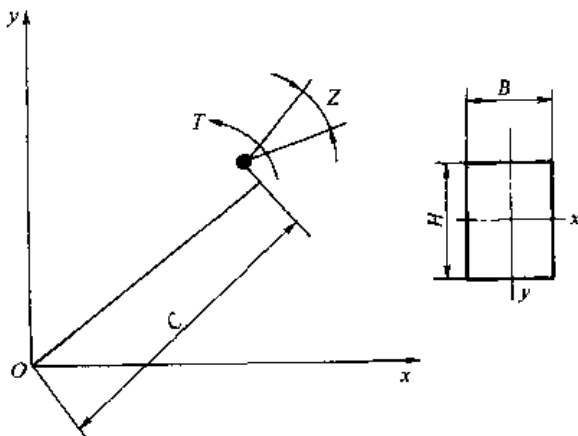


图 44.4-1 柔性梁的变形示意图

根据柔性梁横向振动方程和边界条件,不计结构阻尼,令  $z = \phi(x)q(t)$ ,可得到如下振型函数:

$$\left. \begin{aligned} \phi(x) &= x \\ \phi_1(x) &= -\sin(3.9266x/L) + 0.027875\sinh(3.9266x/L) \\ \phi_2(x) &= \sin \frac{7.0686x}{L} + 1.2041 \times 10^{-3} \sinh \frac{7.0686x}{L} \\ \phi_3(x) &= -\sin \frac{10.2102x}{L} + 5.2032 \times 10^{-5} \sinh \frac{10.2102x}{L} \end{aligned} \right\} \quad (44.4-16)$$

柔性梁在运动过程中的动能和势能分别为

$$T = \frac{1}{2} \int_0^L \rho A \left( \frac{\partial z}{\partial t} \right)^2 dx \quad (44.4-17)$$

$$U = \frac{1}{2} \int_0^L EI_y \left( \frac{\partial^2 z}{\partial x^2} \right)^2 dx \quad (44.4-18)$$

驱动力矩  $T$  所作的功为

$$\delta w = T \sum_0^n \phi'_i(C) \delta q_i \quad (44.4-19)$$

根据拉格朗日方法求出柔性臂在运动过程中的动力学方程如下:

$$M\ddot{q} + Kq = BT \quad (44.4-20)$$

其中

$$M = \begin{bmatrix} m_{00} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & m_{nn} \end{bmatrix}$$

$$K = \begin{bmatrix} k_{00} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k_{nn} \end{bmatrix}$$

$$\ddot{q} = (\ddot{q}_0 \quad \ddot{q}_1 \quad \cdots \quad \ddot{q}_n)^T$$

$$q = (q_0 \quad q_1 \quad \cdots \quad q_n)^T$$

$$B = (\phi'_0(C) \quad \phi'_1(C) \quad \cdots \quad \phi'_n(C))^T$$

其中

$$m_n = \int_0^L \rho A \phi_n^2 dx$$

$$k_n = \int_0^L EI_y (\phi_n'')^2 dx$$

输入扭矩  $T$  和运动端部位移的传递函数为

$$\begin{aligned} G_o(s) &= \frac{Z_L(s)}{T(s)} \\ &= \sum_{i=0}^n \frac{\phi_i(L)\phi'_i(C)}{m_{ii}(s^2 + \omega_i^2)} \\ &= \frac{A_n s^{2n} + A_{n-1} s^{2n-2} + \cdots + A_1 s^2 + A_0}{(s^2 + \omega_0^2)(s^2 + \omega_1^2) \cdots (s^2 + \omega_n^2)} \end{aligned} \quad (44.4-21)$$

其中

$$A_n = \sum_{i=0}^n \frac{\phi_i(L)\phi'_i(C)}{m_{ii}}$$

$$A_{n-1} = \sum_{i=0}^n \frac{\phi_i(L)\phi'_i(C)}{m_{ii}} \sum_{j \neq i} \omega_j^2$$

$$A_{n-2} = \sum_{i=0}^n \frac{\phi_i(L)\phi'_i(C)}{m_{ii}} \sum_{\substack{j, k \neq i \\ j \neq k}} \omega_j^2 \omega_k^2$$

$$\begin{aligned}
 & \vdots \\
 A_0 &= \sum_{i=0}^n \frac{\phi_i(L)\phi_i'(C)}{m_i} \prod_{j \neq i} \omega_j^2 \\
 &= \frac{\phi_0(L)\phi_0'(C)}{m_0} \prod_{j=1}^n \omega_j^2 \\
 Z_L(s) &= \sum_{i=0}^n \phi_i(L)q_i(s)
 \end{aligned}$$

PD 控制的闭环传递函数为

$$G(s) = \frac{(k_p + k_v s) \sum_{i=0}^n A_i s^{2i}}{\prod_{i=0}^n (s^2 + \omega_i^2) + (k_p + k_v s) \sum_{i=0}^n A_i s^{2i}} \quad (44.4-22)$$

其特征方程为

$$\prod_{i=0}^n (s^2 + \omega_i^2) + (k_p + k_v s) \sum_{i=0}^n A_i s^{2i} = 0 \quad (44.4-23)$$

取系统的一阶振型来研究,则

$$G(s) = \frac{(k_p + k_v s)(A_1 s^2 + A_0)}{s^2(s^2 + \omega_1^2) + (k_p + k_v s)(A_1 s^2 + A_0)} \quad (44.4-24)$$

相应的特征方程为

$$s^2(s^2 + \omega_1^2) + (k_p + k_v s)(A_1 s^2 + A_0) = 0 \quad (44.4-25)$$

对于四阶系统来说,当系统具有两对相同的复根时,主导极点离虚轴最远,系统亦具有最佳的稳定性。为了使特征方程式(44.4-25)具有两个重复根,则有

$$k_p = \frac{A_0}{A_1^2} \quad (44.4-26)$$

$$k_v = \frac{2}{A_1} \sqrt{\omega_1^2 - \frac{A_0}{A_1}} \quad (44.4-27)$$

此时特征方程的根为

$$s_{1,2} = -\frac{k_v A_1}{4} \pm \frac{\sqrt{k_v^2 A_1^2 - 16k_p A_1}}{4} \quad (44.4-28)$$

其实部和虚部分别为

$$\text{Real}(\lambda_{\text{evg}}) = -\frac{k_v A_1}{4} = -\frac{1}{2} \sqrt{\omega_1^2 - \frac{A_0}{A_1}} \quad (44.4-29)$$

$$\begin{aligned}
 \text{Image}(\lambda_{\text{evg}}) &= \pm \frac{\sqrt{16k_p A_1 - k_v^2 A_1}}{4} \\
 &= \pm \frac{1}{2} \sqrt{5 \times \frac{A_0}{A_1} - \omega_1^2} \quad (44.4-30)
 \end{aligned}$$

假设单柔性臂机构系统的综合指标为输出超调量  $\sigma_p$ 、调整时间  $t_s$ 。由传递函数式(44.4-24)和特征方程式(44.4-25)知该系统是四阶系统。根据超调量  $\sigma_p$  和调整时间  $t_s$  就可以近似计算出系统的阻尼系数  $\xi$  和

圆频率  $\omega$  为

$$\xi = \frac{|\ln \sigma_p|}{\sqrt{(\ln \sigma_p)^2 + \pi^2}} \quad (44.4-31)$$

$$\omega_T = \frac{3}{t_s \xi} \quad (44.4-32)$$

则其主导极点定义如下:

$$s_{1,2}^* = -\xi \omega_T \pm j \omega_T \sqrt{1 - \xi^2} \quad (44.4-33)$$

对于本系统来说通常可以根据性能指标的要求,在复平面上画出满足这一要求的闭环系统主导极点所在的区域如图 44.4-2 的阴影区域,图中  $\sigma_0 = -\xi \omega_T$  为  $s_{1,2}$  的实部,阻尼角  $\beta = \arccos \xi$ 。

为了满足系统的性能要求,必须使控制系统的主导极点位于图 44.4-2 中的阴影区域。为了使主导极点  $s_{1,2}$  位于图 44.4-2 的阴影区域,由式(44.4-29)和式(44.4-30)可得下式:

$$-\frac{1}{2} \sqrt{\omega_1^2 - \frac{A_0}{A_1}} \leq -\xi \omega_T \quad (44.4-34)$$

$$\omega_1^2 - 5 \frac{A_0}{A_1} < 0 \quad (44.4-35)$$

$$\arctan \left| \frac{\text{Image}(\lambda_{\text{evg}})}{\text{Real}(\lambda_{\text{evg}})} \right| \leq \beta \quad (44.4-36)$$

$$A_1 \geq 0 \quad (44.4-37)$$

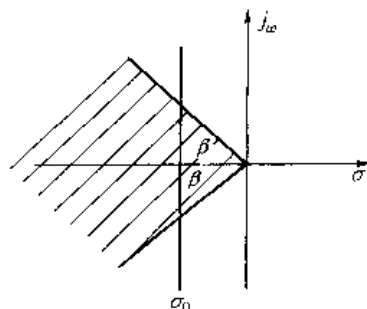


图 44.4-2 闭环极点取值范围示意图

从式(44.4-26)~式(44.4-30)、(44.4-34)~式(44.4-37)可知,  $\omega_1$ 、 $A_0$ 、 $A_1$  以及闭环控制传递函数的极点实部和虚部都和参数  $L$ 、 $H$ 、 $B$ 、 $C$ 、 $k_p$ 、 $k_v$  有关,并且这几个参数之间具有很强的相互耦合性。因此要想使系统具有最佳的综合性能就必须对以上参数进行优化。

对于柔性臂机构来说,要实现系统的高速高精度要求,在满足系统稳定性的前提下,应尽量减轻系统的运动质量,同时为防止高加速启停引起的共振,应使系统具有较高的自振频率。因此柔性臂机构优化模型为如下:

$$\begin{aligned}
 \min F &= 1000 \times (J - 2 \times \pi / \omega_1) \\
 X &= [H \ B \ C]^T
 \end{aligned} \quad (44.4-38)$$

其中,  $f_1 = J = m_{oc} = \frac{\rho H B L^3}{3}$

$$f_2 = \omega_1 = \sqrt{\frac{EI_2}{\rho A} \beta_1^2} = \sqrt{\frac{E}{12\rho} B \left( \frac{3.9266}{L} \right)^2}$$

s. t.

$$g_1 = 2\xi\omega_1 - \sqrt{\omega_1^2 - \frac{A_2}{A_1}} \leq 0$$

$$g_2 = \omega_1^2 - 5 \times \frac{A_2}{A_1} \leq 0$$

$$g_3 = \arctan \sqrt{\frac{5 \times A_0 - \omega_1^2 A_1}{\omega_1^2 A_1 - A_0}} \leq \beta$$

$$g_4 = -\cos \left( 3.9266 \times \frac{C}{L} + 0.027875 \right)$$

$$\times \cosh \left( 3.9266 \times \frac{C}{L} \right) > 0$$

$$g_5 = \delta_v = \frac{3\rho g L^4}{2EH^2} \leq \delta_{v\max}$$

式中  $\delta_v$  —— 柔性臂在垂直方向的变形；

$\delta_{v\max}$  —— 为柔性臂在垂直方向的最大允许变形。

采用序列二次规划法进行优化设计,初始值为: $L = 0.73\text{m}$ ,  $H = 0.019\text{m}$ ,  $B = 0.0032\text{m}$ ,  $C = 0\text{m}$ ,  $\delta_{v\max} = 5\text{mm}$ ,  $\sigma_p = 15\%$ ,  $t = 0.15\text{s}$  计算结果如表 44.4-1 所示。

表 44.4-1 优化设计结果

	参 数	初 始 值	最 终 优 化 值
设计参数	$L/\text{m}$	0.75	0.75
	$H/\text{m}$	0.019	0.0057
	$B/\text{m}$	0.0032	0.0067
	$C/\text{m}$	0	0.3302
	$k_p$	30.8363	349.1
	$k_c$	0.8240	3.6327
性能参数	$J/\text{kg} \cdot \text{m}^2$	0.0206	0.0134
	质量/kg	0.1231	0.0773
	$\omega/\text{rad} \cdot \text{s}^{-1}$	138.36	285.88
	$f/\text{s}^{-1}$	22.02	45.5
	$\sigma_p$	15%	12%
	$t/\text{sec}$	0.15	0.05
	$\delta_v/\text{mm}$	9.72	4.9
	$\delta_{v\max}/\text{mm}$	5	5

## 第5章 结构优化设计

### 1 结构优化设计概述

结构优化设计是将优化技术与有限元分析技术结合起来,设计满足给定的各种要求最佳结构尺寸、形状等的设计手段。因此,结构优化设计在工程设计中得到了广泛的应用,对工程结构设计具有重要意义。结构优化设计突破了传统的结构设计范畴,克服了传统设计经验、类比或采用许多假设和简化导出的计算公式进行结构设计再校核的诸多局限。结构优化设计充分利用了计算机技术、有限元技术和优化技术,可大大缩短设计周期,降低产品材料消耗,提高产品精度和性能,并将产生明显的经济效益。

从广义角度看,结构优化设计包括结构尺寸优化、形状优化、拓扑优化和布局优化;从优化性能角度,结构优化设计包括结构可靠性指标的结构优化、材料性能的结构优化、动力学性能的结构优化和控制结构优化等众多分支。从优化算法方面可以将结构优化方法分为四大类,即优化准则法、数学规划法、优化准则与数学规划结合的混合法及近若干年出现的模拟自然界生物生长和进化的优化算法(如自适应生成法、遗传算法等)。本章仅涉及考虑结构尺寸参数方面的结构优化设计。

与一般优化问题相比,结构优化问题具有变量多、约束多且约束大都为复杂隐式约束的特点。在结构优化问题中,计算位移和应力约束的结构有限元分析耗时长,而优化迭代需多次进行结构有限元分析计算,因此计算效率是结构优化设计中关注的问题。影响结构优化效率的因素是多方面的,如结构优化算法的有效性、结构分析、敏度分析、约束条件的计算效率等等。一方面,有效的结构优化算法迭代次数少,可直接减少计算量;另一方面,采用有效的近似技术,如结构分析的近似、结构性态约束的近似、敏度分析的近似等,也可显著地减少计算量。

在结构优化设计中,适合于优化对象的结构优化算法是关键内容。工程设计中的结构大多为较大规模、大规模甚至超大规模的组合结构问题。不同类型单元之间的组合,使得组合结构的性态响应(如位移、应力、频率等)与设计变量之间的关系十分复杂,许多在单一结构情形下十分有效的结构优化算法无法适用于这些复杂组合结构。另外,在组合结构优化问题中,常常是多种类型的设计变量混合在一起同时进行优化,从而

使问题变得极为复杂。因此,适合于复杂组合结构的优化算法是结构优化需要解决的关键问题。在已有的结构优化算法中,目前已有采用数学规划法成功地解决数百个设计变量的组合结构优化实例。采用优化准则法,可解决设计变量数目与具体问题密切相关。对一般问题可利用射线步的组合结构,采用优化准则法可解决有数千个设计变量的问题,采用混合法也可解决有上千个设计变量的问题。对于简单组合梁,其设计变量数可高达上百万个的惊人程度。当前优化算法的研究多集中在优化准则方面,如板壳结构的优化准则、动力结构的优化准则、基因遗传算法等方面。为了使结构优化软件功能更强大,应针对结构优化问题自身的特点建立结构优化方法程序库。直接采用通用的非线性规划数学方法库,这种方式无法充分利用结构优化问题自身的特点,因而无论是在迭代效率还是在收敛性方面均逊于采用结构优化方法库的方法。采用结构优化方法库的方法,可以对不同的结构优化问题选择不同的结构优化算法,从而可使计算效率和迭代收敛性得到明显提高。另外无论在数学规划还是在优化准则算法迭代中,往往需对一些由具体问题决定的参数进行合理的调整,从而可加速迭代收敛速度或改善迭代稳定性。因此,采用具有自适应策略的优化算法也是十分必要的,如具有频率约束的大规模空间刚架重量优化的自适应优化准则、包含自适应能力的调整约束函数的近似方法以及自适应改变有限元网格的方法。另外,采用人工智能和专家系统技术的结构优化智能化方法,对结构优化的实用化也具有重要意义。

适合于优化对象的结构敏度分析是结构优化设计又一关键内容。现有的有效结构优化算法大多要求使用函数的导数信息,这就需要在结构分析的基础上利用结构敏度分析提供这些导数信息。在结构优化中,结构分析从通用性考虑一般采用有限元法,一些特殊的问题也可采用边界元法。在早期的结构优化软件中,大多采用自行开发的结构有限元分析软件,可处理的单元类型、约束类型、变量类型及结构分析类型均十分有限。自行开发的结构有限元分析软件一般通用性较差,工作量大且为重复性开发工作。由于现已有大量的优秀商用有限元软件可供利用,故一般不再自行开发结构有限元分析软件供结构优化使用。早期商用有限元分析软件由于当时的局限性,并未考虑结构优化的需要,不具备提供导数信息的敏度分析功能;现有商用有

限元分析软件已具备了一定程度的敏度分析功能和结构优化功能,但优化功能仍较弱,因此在商用结构有限元分析软件基础上,开发通用结构敏度分析及优化软件不仅可能,而且也十分必要。在商品化有限元分析程序基础上开发敏度分析功能,通常可由两种方式来实现:一是在有限元分析程序内部加入敏度分析模块;二是将分析软件当作一个黑箱来调用,为外部结构敏度计算提供信息。前一方法实施起来较简单,敏度计算效率高,但对所采用的有限元分析程序的结构、模块功能及内部变量含义、文件管理方式等需要知道得十分清楚,后一方法的优点是不需要分析、消化结构分析软件的内部细节,但实施起来工作量巨大且效率低,尤其对复杂结构和非线性结构更是如此。因此,对于以实用化为目标的结构优化软件,采用在商用有限元分析软件内部开发敏度分析系统的方法更合适。

## 2 结构优化设计的准则法

结构优化准则法,可以认为包括利用力学概念的某些力学准则法,如满应力准则法、应变能密度准则法等和基于数学规划理论中的 Kuhn-Tucker 条件的优化准则法。这类方法的迭代次数与设计变量的多少基本无关,尤其适合应用于大规模结构优化问题。相比之下,数学规划法对于大规模问题几乎不实用,因其计算迭代次数太多。应用结构优化准则法的困难在于:①处理不同类型的结构问题需要导出不同的优化准则;②处理不同性质约束的结构优化问题较为困难,并且有时也难以保证收敛性。

在考虑设计变量为结构单元截面尺寸的尺寸优化问题中,常见的结构重量优化问题的一般数学模型为

$$\text{s. t. } \left. \begin{array}{l} u_{jl} \leq [u_j] \quad j=1, 2, \dots, J; l=1, 2, \dots, L \\ \sigma_{kl} \leq [\sigma_k] \quad k=1, 2, \dots, K; l=1, 2, \dots, L \\ \omega_{pl} \leq [\omega_p] \quad p=1, 2, \dots, P; l=1, 2, \dots, L \\ x_i^l \leq x_i \leq x_i^u \quad i=1, 2, \dots, n \end{array} \right\} \quad (44.5-1)$$

式中  $x$ ——包括杆元面积  $A$ , 板壳膜元厚度  $t$ , 梁元面积  $A$  或惯性矩  $I$ ;

$u_{jl}$ ——位移约束中实际位移值;

$\sigma_{kl}$ ——应力约束中实际应力值;

$\omega_{pl}$ ——频率约束中实际频率值;

$l$ ——工况号。

对式(44.5-1)的结构优化问题求解,可采用多种方法,如数学规划法、优化准则法、混合法、遗传算法等。但是对于不同的问题和不同的求解方法,计算效率上的差异是截然不同的,甚至可能有些方法无法获得结果。因此,对不同的问题应采用适合于该方法

求解,这样才能获得最佳效果。这就需要建立一个方法有效、范围广泛的结构优化算法库以供使用。本节主要讨论针对式(44.5-1)描述的结构优化问题的满应力法等各种准则法。

### 2.1 满应力法

#### (1) 杆结构满应力准则

满应力法通常用于具有应力和尺寸约束的结构优化问题。以杆结构为例,满应力准则是选择各杆面积  $A = [A_1, A_2, \dots, A_n]^T$ , 使各杆应力达到许用应力,即  $\sigma_i = F_i/A_i = [\sigma_i]$ , 其中  $F_i = \max_l \{F_{il}\}$ ,  $F_{il}$  为  $l$  工况下  $i$  杆的内力。对静定结构,满应力设计等价于结构最轻设计,即相当于最优解。并且,此时满应力解即是数学不动点理论中的不动点。对于非静定结构,满应力解就不一定是最轻设计了,并且由于杆的内力已是设计变量的函数,随设计变量的变化而变化,故需多次迭代求解。这时可采用下述迭代公式:

$$\begin{aligned} A_i^{(k+1)} &= \beta_i^{(k)} A_i^{(k)} \\ \beta_i^{(k)} &= \frac{\sigma_i^{(k)}}{[\sigma_i]} \end{aligned} \quad (44.5-2)$$

式中  $k$ ——迭代号;

$A_i$ —— $i$  杆面积;

$\sigma_i$ —— $i$  杆应力;

$[\sigma_i]$ —— $i$  杆许用应力。

事实上式(44.5-2)也可用于静定结构,这时仅需一次迭代即可得到最优解。

#### (2) 板结构满应力准则

在板结构优化问题中,一般使用当量应力(如 von Mises 当量应力)来衡量某一点或某一单元的应力水平。通过有关假设与推导可以得到板结构满应力迭代公式:

$$x_i^{(k+1)} = x_i^{(k)} \left( \frac{\sigma_v}{[\sigma_i]} \right)^{1/2} \quad (44.5-3)$$

式中  $x_i$ ——第  $i$  个板元厚度;

$\sigma_v$ ——第  $i$  个板元当量应力;

$[\sigma_i]$ ——第  $i$  个板元许用应力;

$k$ ——迭代号。

需要指出,这里的满应力迭代公式忽略了板结构单元中的膜应力效应,只考虑了纯弯曲效应。与非静定杆结构类似,这里的迭代公式(44.5-3)对非静定板结构只是一种近似迭代公式,但对静定纯弯曲板结构,该迭代公式是精确的。另外,为了改善非纯弯曲情况下的迭代精度,式(44.5-3)中的  $\sigma_v$  只考虑了膜应力效应和弯曲应力效应的当量应力。

### 2.2 单位移约束准则法

对于只有一个位移约束的问题,由 Kuhn-Tucker

条件容易得到优化准则为

$$\frac{\partial F}{\partial x_i} - \lambda \frac{\partial u}{\partial x_i} = 0 \quad (44.5-4)$$

式中  $F$ ——目标函数;  
 $u$ ——位移值;  
 $\lambda$ ——拉格朗日乘子。

上式对于杆结构有  $\lambda = \frac{W}{[u]}$

式中  $W$ ——结构总重量;  
 $[u]$ ——允许位移值。

从而可得杆结构迭代公式:

$$x_i^{(k+1)} = x_i^{(k)} \left[ \frac{W}{[u]} \frac{1}{\rho_i L_i} \left( -\frac{\partial u}{\partial x_i} \right) \right]^\eta \quad (44.5-5)$$

式中  $x_i$ —— $i$  杆截面积;  
 $\rho_i$ —— $i$  杆材料密度;  
 $L_i$ —— $i$  杆长度;  
 $\eta$ ——松弛因子。

如果采用划分主、被动单元的方法,则有满足 K-T 条件的主动单元的设计变量

$$x_i = \left( \lambda \frac{S_j^k S_j^0}{E_i \rho_i} \right)^{1/2} \quad (44.5-6)$$

$$\lambda = \left[ \frac{\sum_{j=1}^{e_a} \sqrt{S_j^k S_j^0 L_j^2 \rho_j}}{[u] - u_0} \right]^2 \quad (44.5-7)$$

$$u_0 = \sum_{j=1}^{e_p} \frac{S_j^k S_j^0 L_j}{E_j x_j}$$

式中  $S_j^k$ —— $j$  单元在实际载荷下杆的内力;  
 $S_j^0$ —— $j$  单元在单位虚载荷下杆的内力;  
 $e_a$ ——主动单元个数;  
 $e_p$ ——被动单元个数;  
 $E_i$ —— $i$  单元材料的弹性模量。

对于非静定结构式(44.5-5)为一迭代公式,其中  $S_j^k$ 、 $S_j^0$  与杆面积  $x_i$  有关。

### 2.3 多位移约束准则法

对于多位移约束,假定有  $m$  个位移约束在最优点处变成临界约束,则有

$$\frac{\partial F}{\partial x_i} + \sum_{j=1}^m \lambda_j \frac{\partial u_j}{\partial x_i} = 0 \quad (44.5-8)$$

式中  $m$ ——临界位移约束个数;  
 $F$ ——目标函数。

进而迭代公式为

指数迭代式:

$$x_i^{(k+1)} = x_i^{(k)} \left[ \left( \sum_{j=1}^m \lambda_j \frac{\partial u_j}{\partial x_i} \right) / \frac{\partial F}{\partial x_i} \right]^{1/\gamma} \quad (44.5-9)$$

线性迭代式:

$$x_i^{(k+1)} = x_i^{(k)} \left[ \alpha + (1-\alpha) \left( \sum_{j=1}^m \lambda_j \frac{\partial u_j}{\partial x_i} \right) / \frac{\partial F}{\partial x_i} \right] \quad (44.5-10)$$

式中  $\gamma, \alpha$ ——步长参数;  
 $\lambda_j$ ——拉格朗日乘子,  $\lambda_j$  由下式计算:

$$\lambda_j^{(k-1)} = \lambda_j^{(k)} \left( \frac{u_j}{[u_j]} \right)^{1/b} \quad (44.5-11)$$

式中  $b$ ——步长参数。

## 3 结构优化的齿行法

上节介绍的结构优化准则法,通常只能解决具有位移约束的问题(优化准则法)或只具有应力约束的问题(满应力法)。对于应力和位移约束同时存在的问题,本节介绍的齿行法往往十分有效,尤其是对单元刚度与设计变量成线性关系的杆、梁、薄壁结构等更是有效。

结构优化的齿行法主要思想是:根据当前设计点,从所有应力和位移中挑选出一个约束,称之为最严(临界)约束,然后对当前设计点作射线比例变换,使之调整到最严约束面上。而最严约束的选取能够保证经过射线比例调整的点将不违反所有约束条件,即该点为一可行点,随后再通过合理的调整策略修改设计点,使目标函数下降。这样,通过反复迭代可以达到最优。结构优化齿行法的主要优点在于:在迭代过程中,可以视不同情况,无需进行结构分析步,而仅需利用射线步就可及时地将迭代过程中的设计点拉回到约束空间的边界面上,从而保证了迭代能迅速有效地进行。当然,这一方法在非射线步时,仍需结构分析步。

### 3.1 结构优化齿行法简介

结构优化齿行法包括如下三个主要步骤:

- 1) 选择最严约束;
- 2) 走射线步;
- 3) 走改善步。

上述三步的迭代过程(二个设计变量的优化问题)如图 44.5-1 所示。

齿行法无需区分主动变量和被动变量。对于复杂大规模问题,要完成这一任务,是十分困难的。这里需要指出的是,在实际问题中,经常会出现最严约束不是一个,而是含有多个约束的最严约束集,这就需要在最严约束集合中如何选取起主要作用的最严约束。根据我们的研究,这里给出三种选取最严约束的策略供具体问题中应用。

- 1) 优先级策略 优先选取应力约束,而位移约束次之。
- 2) 最严位移约束敏度模控制策略 若最严约束

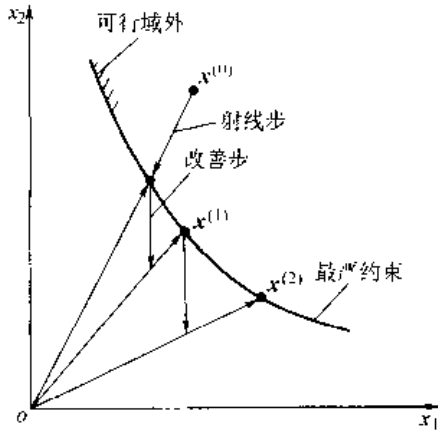


图 44.5.1 两变量齿行线迭代过程示意图

集中,约束均为位移约束,则可按各个位移约束的导数中小于零的分量模之和判断,取模大者为最严约束。

3) 最严应力约束选取策略 若最严约束集合中有两个以上应力约束时,则选取应力梯度较大的应力约束为最严约束。

此外,齿行法迭代过程中松弛因子  $\eta$  的选择对收敛速度及稳定性均有较大的影响。 $\eta$  取小,收敛速度减慢,但稳定性提高;反之, $\eta$  取大,收敛速度增大,但稳定性降低,易出现振荡现象。

### 3.2 杆结构优化齿行法

对于杆结构,齿行法的迭代公式如下:

1) 由当前设计点选取最严约束  $\beta_{max}$ :

$$\beta_{max} = \max_{i,j} (\alpha_i, r_{ij})$$

$$\alpha_i = \sigma_i^{(k-1)} / [\sigma_i]$$

$$r_{ij} = u_j^{(k-1)} / [u_j]$$
(44.5-12)

式中  $\sigma_i$  —  $i$  工况下  $i$  单元的应力;  
 $u_{ij}$  —  $i$  工况下  $j$  自由度方向的位移;  
 $[\sigma_i]$  —  $i$  单元的许用应力;  
 $[u_j]$  —  $j$  自由度方向的允许位移;  
 $k$  — 迭代序号。

若最严约束为一集合(两个约束以上),则按前述策略选取最严约束。

2) 走射线步,取

$$x_i^{(k)} = \beta_{max} x_i^{(k-1)} \quad (44.5-13)$$

使新设计点从某不可行点(或可行点)拉回到最严约束边界面上。

3) 走改善步。若最严约束为应力约束,则使用满应力法修改设计变量,使之满足满应力准则。其迭代公式为

$$x_i^{(k+1)} = x_i^{(k)} \left[ \frac{\max(\sigma_i)^{(k)}}{[\sigma_i]} \right]^{\eta} \quad (44.5-14)$$

若最严约束为位移约束,则使用满位移准则来修改设计变量。

4) 返回步骤 1), 反复迭代。若两次射线步之间目标函数值之差小于给定精度,则认为收敛。

### 3.3 梁结构优化齿行法

在梁结构中,对于平面梁结构,单元刚度矩阵与抗弯惯性矩  $I$  成线性关系。这时采用以  $I$  为设计变量的齿行法就可以取得较好的效果。这里以矩形截面梁单元为例,给出一种新的齿行法。

假设第  $i$  梁单元矩形截面的宽度和高度分别为  $b_i, h_i$ , 则齿行法步骤如下:

1) 选取最严约束:

$$\beta = \max_{i,j} \left( \gamma_j^{(k)}, \alpha_i^{(k)} \right) \left\{ \right. \quad (44.5-15)$$

$$\gamma_j^{(k)} = \frac{u_j^{(k)}}{[u_j]}$$

而  $\alpha_i^{(k)}$  则由对各单元求解下列优化子问题得到

$$\min. \quad (b_i, h_i)$$

$$\text{s. t.} \quad \sigma_{in} / [\sigma_i] - 1, 0 = 0$$

$$b_i^l \leq b_i \leq b_i^u$$

$$h_i^l \leq h_i \leq h_i^u$$
(44.5-16)

式中  $\sigma_{in}$  —  $i$  单元最大当量应力;

$b_i^l, b_i^u$  —  $i$  单元宽度下、上界;

$h_i^l, h_i^u$  —  $i$  单元高度下、上界。

式(44.5-16)优化问题实质上是求单元应力达到许用应力时,使单元重量最小的优化问题,此问题中还可允许包含  $b_i$  与  $h_i$  的关系约束。求解式(44.5-16)描述的优化问题可得  $b_i^l, h_i^l$ , 然后计算

$$a_i^{(k)} = b_i^l (h_i^l)^3 / (b_i^u (h_i^u)^3)$$

2) 走射线步。对各单元求解下列子优化问题:

$$\min. \quad (b_i, h_i)$$

$$\text{s. t.} \quad b_i h_i / [b_i^{(k)} (h_i^{(k)})^3] = \beta$$

$$b_i^l \leq b_i \leq b_i^u$$

$$h_i^l \leq h_i \leq h_i^u$$
(44.5-17)

可得  $b_i^{(k+1)}$  和  $h_i^{(k+1)}$ , 并计算

$$I_i^{(k+1)} = \frac{b_i^{(k+1)} (h_i^{(k+1)})^3}{12} \quad (44.5-18)$$

此时即自动完成了射线步,并能保证走射线步后,单元重量最轻。

3) 走改善步。先求  $I_i^{(k+2)} = \xi^{\eta} I_i^{(k+1)}$ , 若最严约束为位移约束,则取

$$\xi_i = \left[ \frac{W I_i^{(k+2)}}{[u_j] \rho_i L_i A_i^{(k+2)}} \left( - \frac{\partial u_j}{\partial I_i^{(k+2)}} \right) \right] \quad (44.5-19)$$

式中  $W$  — 结构总重量;

$[u_j]$  — 最严约束的许用位移;

$L_i$  — 第  $i$  梁单元长度;

$A_i^{(k+1)}$ ——走射线步后单元  $i$  的截面积;

$\rho_i$ ——第  $i$  梁单元材料密度。

这里的射线步不是靠调整比例因子而是靠求解优化子问题来实现的,  $\partial A_i / \partial A_i^{(k+1)}$  可采用梁结构灵敏度分析方法进行计算。

若最严约束为应力约束, 则取

$$\xi_i = a_i^{(k)} / \beta$$

4) 求得  $I_i^{(k+2)}$  后, 计算公式 (44.5-16) 的  $b_i^{(k+2)}$  及  $h_i^{(k+2)}$ , 然后转步骤 1) 进行迭代直至收敛。

## 4 结构优化的敏度分析技术

### 4.1 敏度分析方法

结构敏度分析是为结构优化提供有关结构约束函数及目标函数的一阶甚至二阶导数信息。结构敏度分析包括有差分法、解析法、解析和差分结合的拟解析法。差分法通用性好, 易于实现, 但计算量大; 解析法实现起来较困难, 但计算效率高; 拟解析法在计算解析敏度困难时可予采用。这里主要介绍解析法。

#### (1) 拟载荷法

考虑线性静力有限元分析的系统方程为

$$KU = F \quad (44.5-20)$$

如果载荷  $F$  与设计变量向量  $X$  无关, 则由式 (44.5-20) 对  $X$  求导可得

$$\frac{\partial K}{\partial X} U + K \frac{\partial U}{\partial X} = 0 \quad (44.5-21)$$

令  $R = -\frac{\partial K}{\partial X} U$ , 则由式 (44.5-21) 可得拟载荷法求结构位移导数的公式为

$$\frac{\partial U}{\partial X} = K^{-1} R \quad (44.5-22)$$

$R$  称为虚拟载荷。有限元法中, 总刚矩阵  $K$  为一稀疏带状对称正定矩阵。一般式 (44.5-20) 采用 Coleskey 三角分解来求解, 亦即相当于  $K^{-1}$  已在结构分析中得到, 从而易于获得  $\partial U / \partial X$ 。这里  $\partial U / \partial X$  的第  $i$  行为

$$\frac{\partial u_i}{\partial X} = \left[ \frac{\partial u_i}{\partial x_1}, \frac{\partial u_i}{\partial x_2}, \dots, \frac{\partial u_i}{\partial x_n} \right]$$

从式 (44.5-22) 可知, 拟载荷法适合于求解所有位移对所有设计变量或对某一个设计变量的导数。

#### (2) 单位载荷法

假想仅在位移  $u_i$  对应的节点和方向上施加单位载荷  $E_i$ , 设其相应的位移响应为  $U^E$ , 由式 (44.5-20) 对  $x_i$  求导可得

$$\frac{\partial U}{\partial x_i} = -K^{-1} \frac{\partial K}{\partial x_i} U \quad (44.5-23)$$

对式 (44.5-23) 两边前乘单位载荷向量  $E_i^T$  可得

$$\frac{\partial u_i}{\partial x_i} = -E_i^T K^{-1} \frac{\partial K}{\partial x_i} U$$

将上式两边转置, 并利用  $K$  的对称性易得单位载荷法求位移导数的公式为

$$\frac{\partial u_i}{\partial x_i} = -U^T \frac{\partial K}{\partial x_i} U^E \quad (44.5-24)$$

此方法适合于求某些位移自由度对所有设计变量或对某些设计变量的导数。一般当设计变量较少、位移约束数目较少时用此方法较经济。

#### (3) 性态空间法

前两种敏度分析方法是在设计空间中进行的。通常称应力、位移等结构性态响应为性态变量, 因此在性态空间中约束可表示为

$$g_j(X, U) \leq 0 \quad j=1, 2, \dots \quad (44.5-25)$$

定义伴随变量向量  $\Phi_j$  使之满足

$$K\Phi_j = \frac{\partial g_j^T}{\partial U} - Z^T$$

由式 (44.5-25) 对  $X$  求导可得

$$\frac{dg_j}{dX} = \frac{\partial g_j}{\partial X} + Z^T \frac{\partial U}{\partial X} \quad (44.5-26)$$

将拟载荷法公式 (44.5-22) 两边左乘  $Z^T$  有

$$Z^T \frac{\partial U}{\partial X} = -Z^T K^{-1} \frac{\partial K}{\partial X} U \quad (44.5-27)$$

将式 (44.5-22) 代入式 (44.5-26) 并利用  $K$  的对称性可得

$$\frac{dg_j}{dX} = \frac{\partial g_j}{\partial X} + \Phi_j^T \left( -\frac{\partial K}{\partial X} \right) U = \frac{\partial g_j}{\partial X} + \Phi_j^T R \quad (44.5-28)$$

当  $g_j = u_i$  时,  $\frac{\partial g_j}{\partial U} = E_i$ ,  $\frac{\partial g_j}{\partial X} = 0$ , 此时式 (44.5-28) 变为

$$\frac{dg_j}{dX} = \frac{du_i}{dX} k = \Phi_j^T R \quad (44.5-29)$$

式 (44.5-29) 为用性态空间法计算位移导数的公式。事实上, 单位载荷法可以从这里的性态空间法或前面的拟载荷法导出。性态空间法不但适用于求位移约束导数, 而且也适用于求应力约束导数。只要对某个约束  $g_j$  求出了伴随变量  $\Phi_j$ , 就可以求得其敏度值。

#### (4) 应力敏度分析

应力敏度分析用来获取应力导数。考虑应力计算有限元公式

$$\sigma = SU \quad (44.5-30)$$

式中  $\sigma$ ——应力;

$S$ ——应力矩阵;

$U$ ——位移向量。

由式 (44.5-30) 对  $X$  求导, 并考虑到尺寸优化时  $S$  与  $X$  无关, 故有应力敏度分析公式为



$$\frac{\partial \sigma_j}{\partial x_i} = S \frac{\partial U}{\partial x_i} \quad (44.5-31)$$

从式(44.5-31)可知,计算应力导数实质上是计算位移导数。由于结构优化中,几乎所有单元应力均有约束,因此应力导数计算量相当大。

考虑到应力约束具有局部特性,故实际应用时可以采用近似方法计算应力导数,其效果是计算量剧减而又能具有足够的精度。这种近似计算方法是将结构在某一迭代步中作暂时静定化处理,冻结结构内力,即近似认为结构内力在某迭代步中与设计变量无关。

对于杆,则有

$$\frac{\partial \sigma_j}{\partial x_i} = \frac{\partial (F_j/x_j)}{\partial x_i}$$

当  $i=j$  时,有

$$\frac{\partial \sigma_j}{\partial x_i} = -\frac{F_j}{x_j^2} = -\frac{\sigma_j}{x_j}$$

当  $i \neq j$  时,有

$$\frac{\partial \sigma_j}{\partial x_i} = 0$$

对于其他类型单元,结构应力约束往往采用 von Mises 当量应力约束。

#### (5) 梁结构位移敏感度分析

在结构优化中,梁结构是十分复杂而且难以处理的问题,尤其对于受弯、扭、剪及轴力作用的空间梁结构更是如此。一方面对梁单元来说,单元刚度矩阵十分复杂,它不仅与单元截面积  $A$  有关,而且还与抗扭惯性矩  $I_t$ 、抗弯惯性矩  $I_2$  和  $I_3$ ,以及剪切面积  $A_2$ 、 $A_3$  等诸多几何尺寸因素有关。另一方面梁单元截面形状种类繁多,不同的截面形状呈现出不同的复杂力学性能。为了减少设计变量数目及简化问题的复杂程度,许多方法往往假设截面特性参数与截面面积之间有函数关系,例如  $I_t = a_1 A^b$ , 这里  $a_1$ 、 $b$  根据不同截面选取不同的值。通过这一函数关系的引入,每个梁单元可只设一个设计变量,如面积或抗弯惯性矩  $I$ ,使敏感度分析及优化过程也变得较为简单。但这种方法的假设与实际工程梁结构差别太大,难以实际应用。这里给出的方法,则可以以单元截面的具体几何尺寸为设计变量。

对于空间梁单元,每个节点有6个自由度,其单元刚度  $K^e$  十分复杂。一般情况下,  $K^e$  为  $A_1$ 、 $A_2$ 、 $A_3$ 、 $I_1$ 、 $I_2$ 、 $I_3$  的函数,而这6个几何特性量与具体截面类型有关,如果直接由单元刚度对具体几何尺寸变量求导,工作量大且求导结果也因截面类型的不同而各异。为了避免各截面类型都对单元刚度分别求导,这里引入了梁结构中间变量即敏感度变量的概念。定义梁单元的敏感度变量为

$$Y = [A_1, A_2, A_3, I_1, I_2, I_3]$$

单元刚度矩阵对敏感度变量的导数为

$$\frac{\partial K^e}{\partial Y_i} = \frac{\partial K^e}{\partial A_i}, \quad i=1, 2, 3$$

$$\frac{\partial K^e}{\partial Y_i} = \frac{\partial K^e}{\partial I_j}, \quad i=4, 5, 6; \quad j=1, 2, 3$$

以单元截面几何尺寸为设计变量时,设计变量集合可用向量  $X = [x_1, x_2, \dots, x_n]^T$  表示,  $n$  为设计变量总数。对敏感度变量的导数,可以通过 Jacobian 变换转变为对设计变量的导数:

$$\left\{ \frac{\partial f}{\partial x_i} \right\} = [J_{ij}] \left\{ \frac{\partial f}{\partial Y_j} \right\}$$

$$[J_{ij}] = \frac{\partial Y_j}{\partial x_i}, \quad i=1, 2, \dots, n; \quad j=1, 2, \dots, m$$

式中  $[J_{ij}]$  — Jacobian 矩阵;

$m$  — 敏感度变量数目。

对空间梁结构  $m=6$ 。当函数  $f$  为梁单元刚度时,则由上式可求得梁单元刚度矩阵对设计变量的导数,进而可由单位虚载荷法或拟载荷法求得位移导数。

## 4.2 敏感度分析的实现

在结构分析软件中增加敏感度分析系统的原则是,敏感度分析系统应作为独立的功能模块,并与原结构分析程序相结合,敏感度分析的模块不影响也不破坏原有程序的结构和功能。

敏感度分析模块需要参与优化的单元单刚的求导信息,为节省计算量,可以直接利用现有结构分析软件的单刚矩阵。这就要求在结构分析计算单刚前,输入必要的控制信息,如哪些单元的单刚要求导,对哪个设计变量求导等。敏感度分析算法还用总刚、位移和应力信息等,这些信息的获取只能在结构分析求解位移线性方程组,并计算出应力、内力等之后。

图 44.5-2 是敏感度分析模块的总体结构,图中未包括结构分析自身的模块。图 44.5-3 和图 44.5-4 分别是拟载荷法和单位虚载荷法的框图。

### (1) 单刚求导数 $\partial K^e / \partial x_i$ 的实施

在有限元计算中,各单元的有关计算多是在局部坐标系中给出的,需转换到统一的总体坐标系中才能组装。因此为提高效率,可以待单刚坐标转换之后对单刚求导。在实施求导过程中,  $\partial K^e / \partial x_i$  中只有一个  $\partial K^e / \partial x_m \neq 0$ , 其余  $\partial K^e / \partial x_i = 0, i=1, 2, \dots, n, i \neq nm$ 。这里  $nm$  是参与优化的单元对应的变量号,  $n$  是设计变量总数。因此,将单刚对全部设计变量求导时,只用计算单刚对一个设计变量的导数。

### (2) 单位应力载荷下的位移响应 $U^E$ 的求解

由于商用软件大多能处理多工况问题,故仅只需将单位虚载荷作为一种载荷工况来处理,即可利用有限元分析程序自身求得  $U^E$ 。

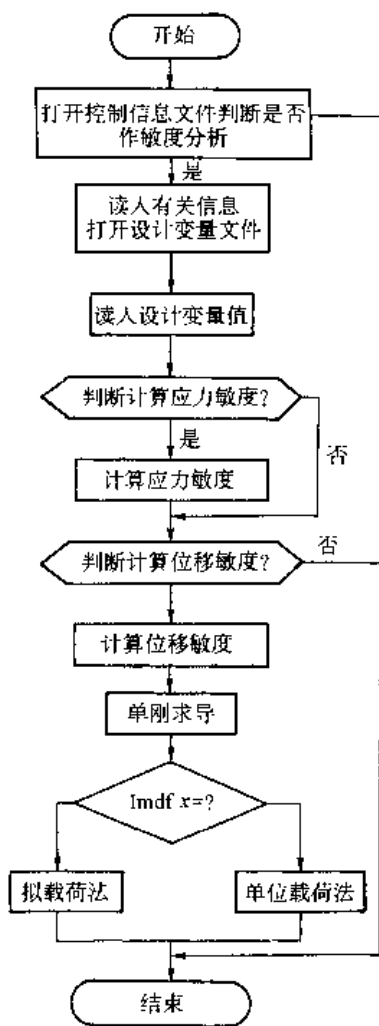


图 44.5.2 敏度分析系统总体结构

(3) 求  $U^T \frac{\partial K}{\partial x_i} U^E$  的实施

$\frac{\partial K}{\partial x_i}$  只有那些与  $x_i$  有关的矩阵元素才为非零值, 因而其非零元素数目较少。在计算  $U^T \frac{\partial K}{\partial x_i} U^E$  时, 只需计算那些非零元素的乘积, 并且不必计算总刚  $\frac{\partial K}{\partial x_i}$ , 而只需直接从  $\frac{\partial K^e}{\partial x_i}$  中求解。这样省去了总刚组集这一庞大的计算工作量, 因而计算效率可大大提高。

(4)  $\frac{\partial K}{\partial x_i}$  的组装

在单位载荷法中, 无需对  $\frac{\partial K^e}{\partial x_i}$  进行组装便可获得  $\frac{\partial K}{\partial x_i}$ 。但是在拟载荷法中进行敏度分析时, 则需要对  $\frac{\partial K^e}{\partial x_i}$  进行组装才能获得  $\frac{\partial K}{\partial x_i}$ , 这是因为在商用有限元软件中直接求  $\frac{\partial K}{\partial x_i}$  是不可能的, 需要通过对  $\frac{\partial K^e}{\partial x_i}$  求导才能间接

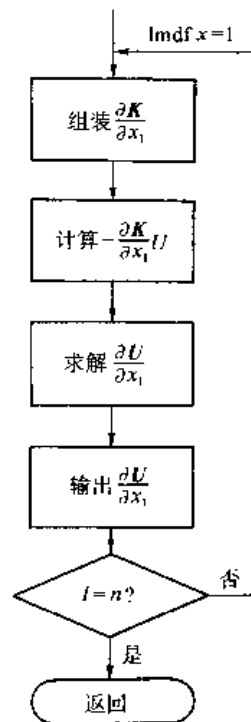


图 44.5.3 拟载荷法敏度分析框图

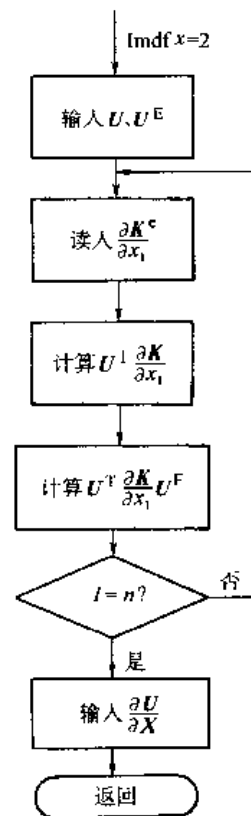


图 44.5.4 单位载荷法敏度分析框图

获得  $\frac{\partial K}{\partial x_i}$ 。需要指出的是:  $\frac{\partial K}{\partial x_i}$  的组装与  $K$  的组装从方法上来说是完全相同的, 故这里不再详述。

(5) 求解  $\frac{\partial U}{\partial x_i} = -\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial x_i} \mathbf{U}$

为了充分利用已有的结构分析信息，对于  $\partial U / \partial x_i$  的求解策略应仔细考虑。在用有限元法求解  $U$  时，一般采用 Coleskey 三角分解法。而求解  $\partial U / \partial x_i$  与求解  $U$  类似，是相当耗时的，其中很大一部分时间消耗在三角分解上，真正回代求解时则是十分迅速的。因此，可以直接利用总刚分解后的信息，进行少量的分解回代即可求解  $\partial U / \partial x_i$ 。

5 实例

本节介绍用梁结构齿行法求解的悬臂梁结构优化问题(如图 44.5-5)，悬臂梁结构优化问题的模型如下式所示：

$$\begin{aligned} \min. \quad & W = \sum_{i=1}^2 b_i h_i L_i \\ \text{s. t.} \quad & \sigma_w \leq [\sigma] \\ & h_i \leq 10b_i \quad i = 1, 2 \end{aligned} \quad (44.5-32)$$

$$\begin{aligned} 0.1 &\leq b_i \leq 100 \\ 0.1 &\leq h_i \leq 100 \\ \delta_z &\leq 0.5 (\text{垂直位移}) \\ \delta_x &\leq 2.0 (\text{垂直位移}) \end{aligned}$$

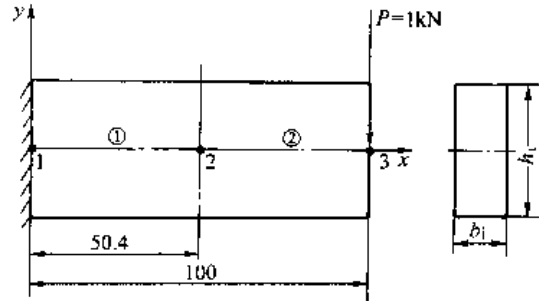


图 44.5-5 悬臂梁优化

采用梁结构齿行法进行优化。模型分为 3 个节点，两个单元， $[\sigma] = 20000\text{N/mm}^2$ ， $E = 10^7\text{N/mm}^2$ 。本问题中长度单位为 m，力的单位为 N。经过 9 次迭代可以得到最优解为  $W_{\min} = 394.17$ ， $b_1 = 0.7107$ ， $h_1 = 7.107$ ， $b_2 = 0.5306$ ， $h_2 = 5.306$ 。

## 第 6 章 形状优化设计

### 1 形状优化设计概述

形状优化是机械和结构优化领域的一个重要分支。在机械和结构设计中,存在最佳过渡曲线形状、结构最佳形状等形状设计问题,因此形状优化有着潜在的应用前景。自从 20 世纪 70 年代初著名学者 Zienkiewicz 提出形状优化问题以来,国内外许多学者在形状优化设计领域开展了大量的研究工作。这些工作主要体现在形状优化的几何建模、形状敏感度分析等方面。

#### (1) 形状优化的几何建模

所谓形状优化的几何建模,即是采用合适的方法来描述待优化的几何形状的边界,而这--边界在优化过程中是不断变化的。几何建模是形状优化要解决的首要问题,因为几何形状的描述方式直接影响着设计变量的表达、结构有限元分析误差、敏感度分析及优化方法等这些形状优化的关键问题。描述边界形状的常用方法有:使用结构离散模型边界节点的坐标作为设计变量,利用这种设计变量的变化来描述边界形状;使用多项式函数来描述边界形状;使用样条函数来描述边界形状以及在微分几何的 Frenet 标架下来描述边界形状等。通过几何建模,可以使可变形形状得到合理的描述,可方便地控制优化边界单元的密度和重划分,同时可减少形状设计变量数目。

#### (2) 形状敏感度分析

形状优化的算法大都需要获得结构状态响应对形状设计变量的导数信息,这就需要进行形状敏感度分析。它涉及到单元刚度矩阵与设计变量之间的复杂非线性关系,因而其敏感度分析及计算相对结构尺寸优化来说要困难得多,并且计算量也要大得多。形状敏感度分析直接影响结构状态响应对形状设计变量导数的数值计算精度和计算效率,而形状优化敏感度分析的计算量往往要占到整个优化过程的一半。因此,提高形状优化敏感度分析的计算效率和计算精度有着重要意义。形状优化敏感度分析方法,从原理上可分为差分法、解析法、拟解析法及变分法等。

#### (3) 有限元网格自动重划分

形状优化迭代过程中,由于设计变量的变化,引起边界形状的相应变化,从而不可避免地引起已划分好的有限元网格也发生变化。如果此时仍然维持原网格的划分方案不变,必然会使网格发生畸变,甚至导

致不合法的网格。因此在形状优化过程中,有限元网格自动重划分是必不可少的。另外,由于形状设计变量的变化,需要对结构某一区域网格进行细化加密,以保证结构有限元分析具有足够的精度。需要指出的是,形状优化过程中,有限元网格自动重划分不是一个简单的网格细化加密问题,而是涉及到完全的重划分。这是由于如果网格划分的方案基本不变,而只是对已有网格加密是无法解决形状优化过程中边界振荡问题的。有限元网格划分及重划分的方法主要有:设计元法、参数映射法、Delaunay 法、二叉树或八叉树法等。

#### (4) 自适应有限元分析

形状优化问题中,形状变量的改变可能会导致有限元网格重划分后,网格产生一定程度的扭曲和畸变,从而大大影响分析精度,甚至使迭代过程失去意义。自适应有限元分析就是通过某种方法,使设计变量发生变化和有限元离散模型发生改变后保证有限元分析精度。目前的自适应分析方法主要可分为  $p$  方法、 $h$  方法、 $r$  方法和  $h-p$  方法。其中, $p$  方法和  $h$  方法更为常用。 $p$  方法是利用提高插值函数阶次的方法来提高有限元分析精度的,而  $h$  方法则是依靠减小单元尺寸来提高分析精度的, $h-p$  方法是将  $h$  方法和  $p$  方法相结合的方法, $r$  方法则是对已有网格靠改变节点位置来改善精度的。 $h$  方法可直接利用现有的结构有限元分析程序来实施,即只需在外围提供有限元网格重划分的结果,不需重新开发分析软件,但需重新开发有限元网格重划分软件。由于要求网格重划分具有高度的自动化和良好的分析精度,使得这一问题的解决具有极大的难度,因而也成了研究的热点之一。 $p$  方法(又称  $p$  单元法)具有高精度和高效率,它对有限元网格重划分的要求甚低,甚至对单元扭曲也不敏感,但是使用时需要重新开发分析软件,这是由于传统的商用有限元软件基本上都不具备  $p$  方法。 $p$  方法很容易与参数化造型系统结合起来,故该软件目前被认为是集成结构形状优化软件的一个较佳选择软件。

#### (5) 自适应误差估计

在自适应分析中需要进行分析误差估计,从而为单元的细划及重划分或插值阶次的提高提供信息。常用的误差估计方法为能量模误差估计,这种估计是基于有限元离散方程的。此方法需要知道精确的结构

应力或位移,但是一般结构的精确值是不可能获得的,只能采用近似方法来近似地确定所谓精确应力或位移。

### (6) 形状优化算法

采用何种优化算法来对形状设计变量进行优化搜索,将直接影响到形状优化的效率。通常优化数值算法有可行方向法、罚函数法、序列线性规划法、序列二次规划法、优化准则法等。可行方向法求解形状优化问题是在每次优化搜索过程中,使用梯度信息寻求一个可行的并且是下降的方向,然后沿该方向作一维搜索,确定合适的搜索步长,获得下一个设计点。这一方法的特点是,每一个新的设计点均是可行设计点,并且优于前一个设计点,因而特别适合于工程优化问题;但是,有时寻找可行且下降的方向是一件相当困难的任务。罚函数法则是通过对原始形状优化问题引入罚项将其转化为一个序列的无约束优化问题,而无约束优化问题相对约束优化问题要容易得多。序列线性规划和序列二次规划法则是分别将原问题转化为一系线性规划或二次规划问题来求解,而线性规划和二次规划的算法是相当成熟的。形状优化准则法目前主要是通过形状优化达到应力或应变能密度沿边界均匀分布或满足其他力学意义的准则,通常用于减少应力集中的形状优化问题。

### (7) 集成形状优化软件

开发一个集成的形状优化软件除了需要解决前述的 6 个关键问题外,还要使整个过程完全自动化。这就需要在几何造型和曲面造型的基础上自动完成几何建模、网格全自动划分与重划分、结构分析及其误差估计等。如果采用自适应形状优化的  $h$  方法来完成这一任务,那么对于几何造型系统与网格全自动划分的耦合是一个主要问题,该问题是一个正在研究的热点。 $h$  方法难度很大, $p$  方法可以避开这一高难度的问题,因此, $p$  方法在集成形状优化软件方面显示出了良好的发展前景。

## 2 形状优化的灵敏度分析

形状优化的灵敏度分析在形状优化问题中占有重要地位。目前大多数形状优化算法都是基于函数的梯度信息的,这靠形状灵敏度分析来实现。形状灵敏度分析主要可分为两类不同的方法:第一类使用离散有限元模型,在有限元系统方程的基础上对形状设计变量直接求导;第二类使用连续质力学中的物质导数法(Material Derivative Method),这种方法是一种变分方法,具有一般性且数值计算精度较高。这两大类方法的差别不在于是否进行结构离散化,而在于何时对结构进行离散。第一类离散模型是先离散后求导,第二类连续模型

是先求导后离散。因此连续模型的物质导数法,结构分析时可采用有限元法、边界元法等数值方法。本节介绍第一类方法,有关第二类方法请参见其他文献。

离子模型形状灵敏度分析又可分为差分法、解析法和拟解析法等。差分法简单,程序易于实现,但计算量太大,一般不予采用;解析法精度高,但程序实现较困难;拟解析法则将某些导数采用差分法计算而另一些导数则采用解析法计算。本节介绍常用的位移和应力灵敏度计算方法。

### 2.1 位移灵敏度计算

对形状优化问题,由有限元系统平衡方法得到

$$KU = F \quad (44.6-1)$$

对上式计算灵敏度有

$$K \frac{\partial U}{\partial x_m} + \frac{\partial K}{\partial x_m} U = \frac{\partial F}{\partial x_m} \quad (44.6-2)$$

如果形状改变的边界处没有外部载荷,则  $\partial F/\partial x_m = 0$ , 此时有

$$K \frac{\partial U}{\partial x_m} = - \frac{\partial K}{\partial x_m} U = R \quad (44.6-3)$$

只要已知  $\partial F/\partial x_m$ , 则可将  $-\partial K/\partial x_m U$  作为拟载荷  $R$  加在结构上,从而可直接利用有限元分析程序求得位移对形状设计变量  $x_m$  的导数。实际问题中,当位移约束只在少数几个节点上存在时,可以利用单位虚载荷法来求解,这样效率更高。现在主要问题在于如何求得精确的  $\partial F/\partial x_m$ , 这对三维形状优化问题并非易事,要花费大量的数值计算时间。下面给出采用拟载荷法进行形状灵敏度分析的方法及步骤。

以三维等参元为例,单元刚度为

$$\begin{aligned} K^e &= \int_{\Omega} B^T D B d\Omega \\ &= \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 B^T D B |J| d\xi d\eta d\zeta \end{aligned} \quad (44.6-4)$$

式中  $B$  —— 单元应变矩阵;

$D$  —— 弹性矩阵;

$|J|$  —— 雅可比变换矩阵。

$D$  与单元形状无关,  $B$  和  $|J|$  则与单元形状有关,故将单元刚度对形状变量  $x_m$  求导可得

$$\begin{aligned} \frac{\partial K^e}{\partial x_m} &= \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \left( \frac{\partial B^T}{\partial x_m} D B |J| + B^T D \frac{\partial B}{\partial x_m} |J| \right. \\ &\quad \left. + B^T D B \frac{\partial |J|}{\partial x_m} \right) d\xi d\eta d\zeta \end{aligned} \quad (44.6-5)$$

上式计算中一般情况下得不到解析解,需使用数值积分方法:

$$\begin{aligned} \frac{\partial K^e}{\partial x_m} &= \sum_{j=1}^{N_G} \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} w_j w_k w_l \left( \frac{\partial B^T}{\partial x_m} D B |J| \right. \\ &\quad \left. + B^T D \frac{\partial B}{\partial x_m} |J| + B^T D B \frac{\partial |J|}{\partial x_m} \right) \end{aligned} \quad (44.6-6)$$

式中  $w_j, w_k, w_l, \dots$ —高斯积分权因子;

$N_G$ —高斯积分阶次。

$$B = [B_1, B_2, L, B_l, L, B_N]$$

$$B_i = \begin{bmatrix} N_{i,x} & 0 & 0 \\ 0 & N_{i,y} & 0 \\ 0 & 0 & N_{i,z} \\ N_{i,y} & N_{i,x} & 0 \\ 0 & N_{i,x} & N_{i,y} \\ N_{i,z} & 0 & N_{i,z} \end{bmatrix}$$

$N_{i,x}, N_{i,y}, N_{i,z}$ 为形函数  $N_i$  对  $x, y, z$  的导数。

$$\begin{bmatrix} N_{i,x} \\ N_{i,y} \\ N_{i,z} \end{bmatrix} = \begin{bmatrix} x_\xi & y_\xi & z_\xi \\ x_\eta & y_\eta & z_\eta \\ x_\zeta & y_\zeta & z_\zeta \end{bmatrix} \begin{bmatrix} N_{i,\xi} \\ N_{i,\eta} \\ N_{i,\zeta} \end{bmatrix} = [J^{-1}] \begin{bmatrix} N_{i,\xi} \\ N_{i,\eta} \\ N_{i,\zeta} \end{bmatrix}$$

偏导数  $x_\xi, y_\xi, z_\xi$  为整体坐标方向  $\xi$  的导数。

$$x_\xi = \frac{\partial x}{\partial \xi} = \sum_{i=1}^n N_{i,\xi} x_i$$

类似地有其他偏导数的计算式。式 (44.6-6) 中:

$$\sigma_i = \frac{1}{\sqrt{2}} \sqrt{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_x - \sigma_z)^2 + 6[(\tau_{xy})^2 + (\tau_{yz})^2 + (\tau_{zx})^2]}$$

式中  $\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{yz}, \tau_{zx}$ —节点  $i$  的六个应力分量。

故当量应力对设计变量的导数为

$$\begin{aligned} \frac{\partial \sigma_i}{\partial x_m} = & \left[ (\sigma_x - \sigma_y) \left( \frac{\partial \sigma_x}{\partial x_m} - \frac{\partial \sigma_y}{\partial x_m} \right) + (\sigma_y - \sigma_z) \left( \frac{\partial \sigma_y}{\partial x_m} - \frac{\partial \sigma_z}{\partial x_m} \right) + (\sigma_x - \sigma_z) \left( \frac{\partial \sigma_x}{\partial x_m} - \frac{\partial \sigma_z}{\partial x_m} \right) \right. \\ & \left. + 6 \left( \tau_{xy} \frac{\partial \tau_{xy}}{\partial x_m} + \tau_{yz} \frac{\partial \tau_{yz}}{\partial x_m} + \tau_{zx} \frac{\partial \tau_{zx}}{\partial x_m} \right) \right] / (2\sigma_i) \end{aligned} \quad (44.6-8)$$

应该指出,对于节点应力分量及其偏导数,不同的单元在单元边界连接处是不连续的,应该使用应力修匀技术。因为一般有限元计算中得到的是单元高斯积分点的应力,所以可以由高斯积分点应力来获得单元节点应力。

$$\{\sigma_i\} = A \{\sigma_i^e\}$$

式中  $\sigma_i^e$ —高斯积分点的应力矩阵;

$A$ —应力修匀矩阵。

$$\text{令: } a = \frac{5+3\sqrt{3}}{4}, b = \frac{\sqrt{3}+1}{4}, c = \frac{\sqrt{3}-1}{4}, d =$$

$$\frac{5-3\sqrt{3}}{4}, \text{ 则 } A \text{ 可表示为}$$

$$A = \begin{bmatrix} a & b & c & b & b & c & d & c \\ & a & b & c & c & b & c & d \\ & & a & b & c & c & b & c \\ & & & a & c & d & c & b \\ & & & & a & b & c & b \\ & & & & & a & b & c \\ & & & & & & a & b \\ & & & & & & & a \end{bmatrix} \quad (44.6-9)$$

$$\frac{\partial B}{\partial x_m} = \begin{bmatrix} \frac{\partial N_{i,x}}{\partial x_m} & 0 & 0 \\ 0 & \frac{\partial N_{i,y}}{\partial x_m} & 0 \\ 0 & 0 & \frac{\partial N_{i,z}}{\partial x_m} \\ \frac{\partial N_{i,y}}{\partial x_m} & \frac{\partial N_{i,x}}{\partial x_m} & 0 \\ 0 & \frac{\partial N_{i,x}}{\partial x_m} & \frac{\partial N_{i,y}}{\partial x_m} \\ \frac{\partial N_{i,x}}{\partial x_m} & 0 & \frac{\partial N_{i,z}}{\partial x_m} \end{bmatrix} \quad (44.6-7)$$

$$\text{式中 } \frac{\partial N}{\partial x_m} = [J^{-1}] \frac{\partial J}{\partial x_m} N$$

利用式 (44.6-6) 和式 (44.6-7) 可求得  $\partial K^e / \partial x_m$ , 从而可得  $\partial U / \partial x_m$ 。

## 2.2 应力敏感度计算

考虑节点当量应力有

因此可得节点应力导数为

$$\frac{\partial \{\sigma_i\}}{\partial x_m} = A \frac{\partial \{\sigma_i^e\}}{\partial x_m}$$

由于  $\sigma_i^e = DBU^e$ , 故

$$\frac{\partial \{\sigma_i\}}{\partial x_m} = D \left( B \frac{\partial U^e}{\partial x_m} + \frac{\partial B}{\partial x_m} U^e \right)$$

式中  $i$ —上标, 表示第  $i$  单元;

$\partial B / \partial x_m$ —偏导数。

由于单元在同样节点处的修匀应力不连续, 故应按式 (44.6-8) 计算光滑修匀节点应力导数:

$$\frac{\partial \{\sigma_i^j\}}{\partial x_m} = \frac{1}{N_i^j} \sum_{k=1}^{N_i^j} \frac{\partial \{\sigma_k^e\}}{\partial x_m} \quad (44.6-10)$$

$$\{\sigma_i^j\} = \frac{1}{N_i^j} \sum_{k=1}^{N_i^j} \{\sigma_k^e\}$$

式中  $\{\sigma_i^j\}$ —节点  $j$  处的光滑应力;

$\{\sigma_k^e\}$ —与节点  $j$  有关的  $k$  单元节点应力;

$N_i^j$ —与节点  $j$  相关的单元数目。

## 2.3 形状敏感度分析实现

形状敏感度分析实现的主要步骤包括:

1) 在第  $k$  次迭代中, 根据边界节点产生结构区域内的节点坐标及网格。边界节点由当前形状设计变量(主控制点)来描述。

2) 用有限元法进行结构分析, 计算结构状态响应。由形状插值函数公式利用  $N_{i,s}$ 、 $N_{i,q}$ 、 $N_{i,r}$  可求得总体坐标对局部坐标的偏导数, 计算  $[J]$ 、 $[J^{-1}]$  和  $[B]$ 。然后在每个高斯积分点处根据当前的节点坐标, 求出  $N_{i,s}$ 、 $N_{i,q}$  和  $N_{i,r}$  构造几何矩阵  $B$  和单元刚度矩阵

$$\frac{\partial K^e}{\partial x_m} \{U_e\} = \sum_{j=1}^{N_G} \sum_{k=1}^{N_G} \sum_{i=1}^{N_G} w_j w_k w_i \left[ \left( \frac{\partial B^T}{\partial x_m} |J| + B^T \frac{\partial J}{\partial x_m} \right) \{\sigma'_e\} + B^T D \frac{\partial B}{\partial x_m} |J| \{U_e\} \right]_{j,k,i}$$

5) 根据每个单元的  $\frac{\partial K^e}{\partial x_m} \{U_e\}$  总装组集得到虚拟载荷向量  $R$ 。

6) 将  $R$  作为虚拟载荷加在原结构上, 求解平衡方程式得位移敏度  $\frac{\partial U}{\partial x_m}$

7) 采用应力修匀技术, 计算节点应力对设计变量的导数  $\frac{\partial \{\sigma'_e\}}{\partial x_m}$ , 然后求得节点处当量应力对设计变量的导数。

### 3 形状优化的自适应分析技术

形状优化中通常采用的结构分析方法为有限元法, 有时也采用边界元法。用有限元法分析结构响应如应力、位移、频率等都是具有计算误差的, 而这种误差与有限元网格剖分密度、网格形状质量、有限元插值函数的阶次等是密切相关的。形状优化问题从早期发展至今已主要成为了自适应形状优化问题, 这主要涉及自适应有限元分析。自适应有限元分析的核心内容为分析误差估计和自适应有限元网格的自动生成。自适应网格自动生成需要根据分析误差的分布状况来自动调整被分析结构各处的网格剖分密度和网格形状, 从而达到要求的分析精度。因此自适应网格自动生成比普通网格自动生成具有更大的难度。在各类网格自动生成算法中较适合于自适应网格生成的有 Delaunay 法和四叉树(或八叉树)法。

自适应分析的目的在于提高分析精度, 以满足各种应用场合的需要。这可有几种途径来实现: 其一为  $h$  方法, 即使用较好形状和合理密度的有限元网格; 其二为  $p$  方法, 即采用高阶插值形函数来获得较高的计算精度; 其三为  $r$  方法, 这一方法靠改变已有网格的节点位置来改善分析精度而不增加单元和节点的数目; 其四为  $h-p$  方法, 即将  $h$  方法和  $p$  方法结合起来。不同的自适应方法又往往采用不同的误差估计器, 但通常都属于后继误差估计器, 即先进行有限元分析, 再进行误

$K^e$  并总装组集总刚阵  $K$ 。再由结构有限元系统平衡方程, 求出节点位移向量  $U$ 。计算过程中的  $K^e$ 、 $U$  及各高斯积分点处的几何矩阵和应力向量, 均存储起来以备后用。

3) 计算  $\frac{\partial [J]}{\partial x_m}$  和  $\frac{\partial N_{i,s}}{\partial x_m}$  等, 并由此得到  $\frac{\partial B}{\partial x_m}$  和  $\frac{\partial |J|}{\partial x_m}$ 。

4) 计算  $\frac{\partial K^e}{\partial x_m} \{U_e\}$ , 再将该式两边同乘以单元节点位移  $\{U_e\}$ , 并考虑到  $\{\sigma'_e\} = DB \{U_e\}$  得

差估计。本节介绍常用的  $h$  自适应方法等。

#### 3.1 误差估计

在误差估计方面, 使用最多且利用现有有限元分析软件较易实现的是 Zierkiewicz 提出的能量模误差估计方法。

考虑一般位移有限元法系统平衡方程

$$K \tilde{u} = F \quad (44.6-11)$$

式中  $K$  ——总刚阵;

$\tilde{u}$  ——有限元位移解;

$F$  ——结构外载荷矢量。

上述方程实际上是下述微分方程的离散近似解:

$$Lu = f \quad (44.6-12)$$

式中  $L$  ——微分算子;

$u$  ——位移精确解;

$f$  ——右端矢量。

定义局部误差向量:

$$e = u - \tilde{u} \quad (44.6-13)$$

式中  $\tilde{u}$  ——网格尺寸为  $h$  时的位移近似解;

$e$  ——局部误差向量。

为了便于数值计算, 一个更常用的误差定义方式为能量模定义:

$$\begin{aligned} \|e\| &= \left[ \int_{\Omega} (Se)^T D (Se) d\Omega \right]^{\frac{1}{2}} \\ &= \left( \int_{\Omega} e^T S D S e d\Omega \right)^{\frac{1}{2}} \\ &= \left( \int_{\Omega} e_o^T D^{-1} e_o d\Omega \right)^{\frac{1}{2}} \end{aligned} \quad (44.6-14)$$

式中  $S$  ——结构应力矩阵;

$D$  ——结构弹性矩阵;

$e_o$  ——应力误差。

$e_o$  与  $e$  类似地定义为

$$e_o = \sigma - \sigma^e \quad (44.6-15)$$

由于式(44.6-14)中的SDS矩阵为自伴的,故有

$$\|e\|^2 = \|u\|^2 - \|\tilde{u}^h\|^2 \quad (44.6-16)$$

$\|u\|$ 称为位移模。假定 $h$ 网格方案进一步细化后的解为 $u^*$ 和 $\sigma^*$ ,则有能量模误差估计式为

$$\|e\| \approx \left( \int_{\Omega} (\epsilon^* - \epsilon^h)^T D (\epsilon^* - \epsilon^h) d\Omega \right)^{\frac{1}{2}} \quad (44.6-17)$$

如果利用应力应变关系:

$$\sigma = D\epsilon$$

则有另一形式的 $\|\tilde{e}\|$ 为

$$\begin{aligned} \|e\| &\approx \|\tilde{e}\| \\ &= \left( \int_{\Omega} (S(u^* - u^h))^T DS(u^* - u^h) d\Omega \right)^{\frac{1}{2}} \\ &= \left( \int_{\Omega} (\sigma^* - \sigma^h)^T D^{-1} (\sigma^* - \sigma^h) d\Omega \right)^{\frac{1}{2}} \end{aligned} \quad (44.6-18)$$

现在,有限元分析误差计算变成了 $\sigma^*$ 和 $\sigma^h$ 的计算问题了。对于式(44.6-18)的计算可有两种方法:第一种方法是在某一网格划分方案( $h$ 方案)下进行有限元分析得到 $\sigma^h$ ,然后将单元剖分成很小尺寸再进行有限元分析得到 $\sigma^*$ ,从而可计算出 $\|e\|$ ;第二种方法则可避免两次有限元计算,即采用应力修匀方法而获得的应力可近似作为 $\sigma^*$ ,这种方法是先利用有限元分析获得节点应力、高斯点应力,然后利用应力修匀方法获得修匀后的节点应力并将其作为 $\sigma^*$ 。

### 3.2 $h$ 自适应法

$h$ 自适应方法通过对误差较大的单元和区域进行网格细化来提高计算精度。采用 $h$ 方法的关键在于通过误差估计决定网格细化的尺寸,因此 $h$ 自适应形状优化的一个重要任务在于形成一个可对网格划分进行灵活控制的全自动网格生成器和一个高效、高精度的误差估计器。 $h$ 自适应方法可直接使用已有的商用结构分析软件,接上用户自己的网格生成器、误差估计器和优化器,即可进行 $h$ 自适应形状优化。

$h$ 方法需要确定网格在不同区域的分布密度和细化尺寸。一般采用的是用自适应单元尺寸来控制单元分布密度,其原则是使每个单元能量模误差相等。为此定义能量模误差平均百分比为

$$\eta = \frac{\|e\|}{\|u\|} = \frac{\|e\|}{(\|u^h\|^2 + \|e\|^2)^{1/2}} \quad (44.6-19)$$

假定 $\eta$ 的允许值为 $\bar{\eta}$ ,对单元 $i$ 有能量模误差比为

$$\xi_i = \frac{\|e\|_i}{\|e\|_p} \quad (44.6-20)$$

式中  $\|e\|_i$ —— $i$ 单元能量模误差;

$\|e\|_p$ ——允许单元误差。

$$\|e\|_p = \bar{\eta} [(\|u^h\|)^2 + (\|e\|)^2]^{1/2} / \sqrt{M} \quad (44.6-21)$$

故有 $\|e\|_i \leq \|e\|_p$

$$= \bar{\eta} [(\|u^h\|)^2 + (\|e\|)^2]^{1/2} / \sqrt{M} \quad (44.6-22)$$

式中  $M$ ——单元总数。

而近似能量模误差 $\|\tilde{e}\|$ 由单元组成,即

$$\|\tilde{e}\|^2 = \sum_{i=1}^M (\|e\|_i)^2 \quad (44.6-23)$$

当 $\xi_i > 1$ 时, $i$ 单元误差超过许用值,单元 $i$ 需要细划。而对有限元分析,下式成立:

$$\|e\| \propto h^p \quad (44.6-24)$$

式中  $p$ ——位移插值函数的阶次;

$h$ ——单元尺寸。

故新单元尺寸可由 $\xi$ 得到

$$h_{i_{\text{new}}} = h_{i_{\text{old}}} / \xi_i^{1/p} \quad (44.6-25)$$

在实际应用时,有时也采用 $L_2$ 范数来进行误差估计,有关此误差估计器的公式如下:

$$\|u\|_{L_2} = \left( \int_{\Omega} u^T u d\Omega \right)^{1/2} \quad (44.6-26)$$

$$\|e_e\|_{L_2} = \left( \int_{\Omega_e} (\sigma - \sigma^h)^T (\sigma - \sigma^h) d\Omega \right)^{1/2} \quad (44.6-27)$$

$$\Delta\sigma = \left( \frac{\|e_e\|_{L_2}^2}{\Omega_e} \right)^{1/2} \quad (44.6-28)$$

式中  $L_2$ ——下标,表示二范数;

$\Omega_e$ ——单元 $e$ 的区域体积(面积);

$\Delta\sigma$ —— $\Omega_e$ 上的局部应力误差。

显然,此误差估计器计算较为简单。

为了描述误差的相对值,通常引入效率指数 $\theta$ 。

$$\theta = \frac{\|e\|}{\|e^*\|} \quad (44.6-29)$$

式中  $\|e\|$ ——估计误差模;

$\|e^*\|$ ——精确误差值。

对实际应用问题,自适应网格加密可进行1~2次。另外对于具有奇异性的问题,如应力集中问题等,可采用 $\Delta\sigma$ 误差估计器,此时效果更好。图44.6-1为采用 $h$ 方法进行自适应分析的某方案流程图。

### 3.3 $p$ 自适应法

$p$ 自适应方法是靠提高插值函数的阶次来提高分析精度。与 $h$ 方法相比, $p$ 方法有许多优点。

在光滑解方面, $p$ 方法和 $h$ 方法的插值函数阶次可以趋于无穷,且解收敛于精确解。 $p$ 方法在所有单元误差趋于相同时即在误差均衡网格情形下,能量模误



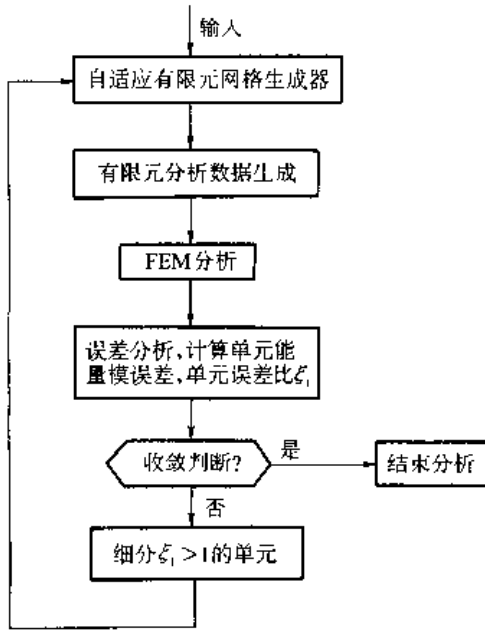


图 44.6-1 h 自适应有限元分析流程图

差满足

$$\|e\| \leq \frac{k_2}{N^{p/2}} \quad (44.6-30)$$

式中  $N$ ——自由度数目;  
 $p$ ——插值函数阶次;  
 $k_2$ ——常数。

对非光滑问题,  $p$  方法的收敛速率至少与  $h$  方法相同; 而对具有角点奇异性 (应力集中) 问题,  $h$  方法的收敛速率是  $p$  方法的两倍。  $p$  方法和  $h$  方法都可以采用优化网格的方式来提高收敛速率, 但  $p$  方法受网格性能的影响远比  $h$  方法要小得多。在网格生成方面,  $p$  方法要简单得多,  $p$  方法可以采用大尺寸、少数目的网格, 故输入数据量也小得多。在矩阵稀疏性方面,  $h$  方法的刚阵比  $p$  方法的要稀疏得多。因此对  $p$  方法应设法尽量提高刚阵的稀疏程度, 以加速计算。在自适应方面,  $p$  方法的自适应技术远比  $h$  方法简单。对于  $p$  方法, 只需加入新的形函数, 无需改变网格。  $p$  方法形函数的增加与  $h$  方法类似, 依赖于能量模误差估计。对能量模变化较大的区域, 可增加新的形函数来提高精度。相比之下,  $h$  自适应方法, 在网格全自动生成和密度控制以及数据管理等方面面临的困难较大, 而将  $p$  方法和  $h$  方法结合起来的  $h-p$  方法面临的难度更大。在误差精度估计方面,  $p$  方法允许用户对感兴趣区域内的数据进行控制, 这种控制基于增加  $p$  自由度 (即插值函数阶次) 后计算数据的比较, 因此用户可事先指定要求的精度。在形函数方面, 在  $p$  方法中形函数具有层次叠加性 (hierarchical function), 而  $h$  方法的形函数为等参单元形函数。对等参单元形函数, 当增加高阶

形函数时, 低阶形函数也需改变, 同时原有的单元刚阵元素全部改变, 必须重新形成刚阵元素, 并且由于形函数阶次增高, 等参单元的节点自由度增加, 其相邻单元的节点自由度也均需增加, 否则需采用过渡单元。在  $p$  方法中的层叠形函数不受这些条件限制, 形函数阶次增加时, 只需在原形函数的基础上叠加上新增加的项, 这类似于函数的傅里叶级数展开项。这类层叠形函数可以按需要来增加高阶项, 不需过渡单元来保证单元间的连续性, 甚至对同一个节点的不同坐标分量, 也可以采用不同阶次的形函数。在算法实施方面,  $h$  方法可以充分利用已有的结构有限元商用软件和网格全自动生成器,  $p$  方法则需另行开发有关的  $p$  单元有限元分析软件。但目前国外已出现了采用  $p$  方法的商用软件, 如三维  $p$  自适应的结构分析软件 STRIPE, 采用  $p$  单元的结构分析软件 MSC/PROBE 和采用  $h-p$  自适应的 PHLEX 软件。而 Rasna 公司的 Applied Structure 软件可以进行  $p$  方法的形状优化。

$p$  方法中的一个关键问题是层叠形函数的定义, 它远比  $h$  方法的等参形函数要复杂得多。

定义  $n$  个基函数  $N_i(\xi, \eta)$ ,  $i=1, 2, \dots, n$ , 则层叠形函数  $u$  可用下述函数表示:

$$u = \sum_{i=1}^n a_i N_i(\xi, \eta) \quad (44.6-31)$$

式中  $a_i$ ——任意实数。

由于  $a_i$  可以为任意实数, 因而式 (44.6-31) 的表示与  $h$  有限元中是不同的。  $h$  有限元中,  $a_i$  需满足单元边界的连续性条件。  $p$  方法对式 (44.6-31) 中的形函数的基本要求为:

- 1) 可分层次叠加计算。当增加高阶自由度时, 不需要修改低阶形函数。
- 2) 阶次越高, 该项对近似解的贡献重要性反而越小。
- 3) 用于有限元时, 节点变量不一定指节点位移, 也可为其他变量, 故而不要求层叠形函数满足下面一般形函数 (等参) 的要求:

$$N_i(\xi_j) = \begin{cases} 1 & \text{如果 } i = j \\ 0 & \text{如果 } i \neq j \end{cases}$$

$p$  方法中, 层叠函数的高阶项的增加, 需由误差估计来决定, 仍采用能量模误差估计。根据局部误差定义有

$$(\|e\|)^2 = \int_{\Omega} e r d\Omega - \int_{\Gamma} e r_i d\Gamma \quad (44.6-32)$$

式中  $r$ ——应力残差;  
 $r_i$ ——表面力残差;  
 $\Omega$ ——求解区域;  
 $\Gamma$ —— $\Omega$  的边界。

假定已采用  $p$  单元获得了有限元解, 节点变量为  $n$  个, 即

$$u_n = \sum_{i=1}^n a_i^{(n)} N_i \quad (44.6-33)$$

$a_i^{(n)}$  满足下列平衡方程:

$$K_{(n)} a^{(n)} = f_{(n)} \quad (44.6-34)$$

如果在上面  $n$  个自由度解的基础上增加  $m$  个自由度后, 节点变量增至  $(n+m)$  个, 则用  $a^{(n+m)}$  表示, 它同样应满足平衡方程, 即有

$$K_{(n+m)} a^{(n+m)} = f_{(n+m)} \quad (44.6-35)$$

$$u_{(n+m)} = \sum_{i=1}^{n+m} a_i N_i = \sum_{i=1}^n a_i N_i + \sum_{i=n+1}^{n+m} a_i N_i$$

利用  $N_i$  的层次叠加性, 则此时能量模误差变化为

$$e_{n,m} = u_{n+m} - u_n = \sum_{i=1}^n (a_i - a_i^{(n)}) N_i + \sum_{i=n+1}^{n+m} a_i N_i$$

显然  $e_{n,m}$  即为实际误差。若近似认为

$$a_i \approx a_i^{(n)} \quad i = 1, 2, \dots, n$$

则有

$$e_{n,m} \approx \sum_{i=n+1}^{n+m} a_i N_i = N^{(m)} a^{(m)}$$

对于层叠函数, 当增加高阶自由度时, 其低阶刚度矩阵元素保持不变, 故式 (44.6-35) 可写为

$$\begin{bmatrix} K_{(n)} & K_{(n,m)} \\ K_{(m,n)} & K_{(m)} \end{bmatrix} \begin{bmatrix} a^{(n)} \\ a^{(m)} \end{bmatrix} = \begin{bmatrix} f_{(n)} \\ f_{(m)} \end{bmatrix} \quad (44.6-36)$$

利用式 (44.6-34) 有

$$a^{(n)} = K_{(n)}^{-1} (f_{(n)} - K_{(n,m)} a^{(m)}) \quad (44.6-37)$$

如果每次只增加一个自由度, 则有

$$e_{n,1} = N_{n+1} a_{n+1}$$

$$a_{n+1} = \frac{1}{k_{n+1,n+1}} (f_{n+1} - K_{n+1,n} a^{(n)})$$

式中  $a_{n+1}$  ——  $a^{(n+1)}$  的第  $n+1$  个分量;

$f_{n+1}$  ——  $f^{(n+1)}$  的第  $n+1$  个分量;

$k_{n+1,n+1}$  ——  $K_{(1,n)}$  的第  $n+1$  个主对角元素;

$K_{n+1,n}$  —— 第  $n+1$  行的前  $n$  个元素。

故有

$$e_{n,1} = \frac{(f_{n+1} - K_{n+1,n} a^{(n)})}{k_{n+1,n+1}} N_{n+1}$$

式中  $N_{n+1}$  —— 新增形函数项。

由于  $k_{n+1,n+1} = (\|N_{n+1}\|)^2$ , 故有

$$(\|e_{n,1}\|)^2 = \frac{(f_{n+1} - K_{n+1,n} a^{(n)})^2}{k_{n+1,n+1}}$$

取误差指数为  $\eta_{n+1}$ , 则

$$\eta_{n+1} = \frac{\left[ \int_{\Omega} N_{n+1}^2 r d\Omega \right]^{1/2}}{k_{n+1,n+1}}$$

对每个增加的高阶自由度  $i$  求出  $\eta_{i-1}$ , 并设为  $\eta_i$ , 令

$$\eta_{\max} = \max(\eta_i)$$

$$a_i \approx \eta_i / \eta_{\max}$$

如果  $a_i \geq r$  则在第  $i$  个自由度上加上高阶项,  $r$  通常取为  $r=0.25; 0.5$ , 利用柯西-许瓦兹不等式并考虑单元跨交界区域内的误差, 则有误差为

$$(\|e_{n,i}\|)^2 = \frac{\left[ \sum_{\Omega} \left[ \int_{\Omega} N_{n+1}^2 r d\Omega \right] \left[ \int_{\Omega} r^2 d\Omega \right] \right]}{\left[ \sum_{\Omega} \left[ \int_{\Omega} N_{n+1}^2 d\Gamma \right] \left[ \int_{\Omega} J^2 d\Gamma \right] \right] / k_{n+1,n+1}} \quad (44.6-38)$$

再由式 (44.6-38), 利用 Zienkiewicz 公式有

$$(\|e_{n,i}\|)^2 = \sum_{\Omega} \left[ \frac{h^2}{24c\rho^2} \int_{\Omega} r^2 d\Omega + \frac{L}{24c\rho} \int_{\Omega} J^2 d\Gamma \right] \quad (44.6-39)$$

#### 4 自适应分析形状优化设计

利用自适应分析即可进行形状优化设计。形状优化设计的常用数学模型为

$$\begin{aligned} \min. & \quad \left\{ \max_{x \in \Omega} |F| \right\} \quad (44.6-40) \\ \text{s. t.} & \quad \int_{\Omega} d\Omega - \bar{A} \leq 0 \end{aligned}$$

式中  $F$  —— 局部位移, von Mises 应力或其他类型应力 (如剪应力等) 之一;

$A$  —— 给定的面积、体积或重量的上界;

$\Omega$  —— 设计区域;

$X$  —— 形状设计变量。

该模型可转化为便于数值计算的下述模型:

$$\min_{\Gamma_d} (\beta) \quad (44.6-41)$$

$$\text{s. t.} \quad \int_{\Omega} d\Omega - \bar{A} \leq 0 \quad |F| - \beta \leq 0 \text{ (在 } \Omega \text{ 上)}$$

式中  $\Gamma_d$  —— 设计区域的边界。

下式是另一种常用的形状优化设计模型:

$$\min_{\Gamma_d} (\alpha) \quad (44.6-42)$$

$$\text{s. t.} \quad |F| - \beta \leq 0 \text{ (在 } \Gamma_d \text{ 上)}$$

式中  $\alpha$  —— 面积、体积或重量。

形状优化设计变量通常均为连续变量, 因而几乎所有的连续变量的数学规划算法都可应用于形状优化。但相对来说, 比较有效的优化算法是序列二次规划、序列线性规划以及内点法等。设计变量的选取一般是参数曲线 (面) 中的控制点 (又称为主节点), 控制点将定义一些超单元, 这样通常可以取较少数目的设计变量即可满足形状控制的要求。通过形状敏度分析提供约束及目标函数的导数信息, 即可直接采用各种数学规划算法进行优化搜索。在形状优化设计中, 自适应重分析通常只进行一次即可认为收敛到了指定分析精度, 以避免过大的自适应分析代价。

图 44.6-2 是采用有限元分析的  $h$  自适应形状优化的流程图。

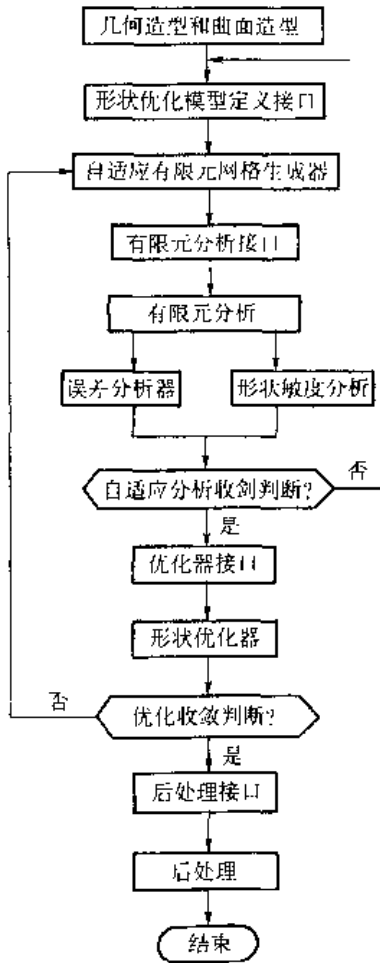


图 44.6-2  $h$  自适应形状优化流程图

### 5 形状优化设计实例

自形状优化设计问题提出以来,已有大量形状优化应用实例见诸各种文献资料。本节给出有关连杆和起重吊钩的二个形状优化实例。

#### 5.1 连杆的形状优化设计

连杆是内燃机关键零件。在进行连杆设计时,希望能得到一个重量轻且安全可靠的连杆结构尺寸。这依靠传统设计方法是无法得到最优设计结果的,采用形状优化设计则可满足上述设计要求。图 44.6-3 为内燃机连杆,主要尺寸如图 44.6-3 所示。

内燃机连杆形状优化设计主要过程如下:

##### (1) 有限元模型建立

为了简化计算模型,将连杆小端的轴孔固定,大端轴孔承受最大分布载荷为 60MPa,且在轴孔的接触角  $\phi=120^\circ$  的区域按余弦规律分布。在进行连杆有限元网格划分时,考虑到连杆各部分厚度不同,可将模型简

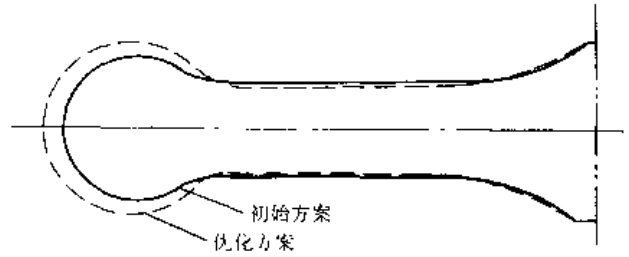


图 44.6-3 连杆优化前、后的形状  
化为变厚度的二维平面模型,共划分为 210 个单元,278 个节点。

##### (2) 形状优化数学模型建立

1) 目标函数 优化目标取为连杆的重量最轻,故有重量函数为

$$W[X] = \sum_{i=1}^{210} \rho_i S_i t_i \quad (44.6-43)$$

式中  $\rho_i$ —— $i$  单元的材料密度;

$S_i$ —— $i$  单元的面积;

$t_i$ —— $i$  单元的厚度。

2) 设计变量 连杆工作时,杆身容易断裂,因此设计变量的选择要能描述连杆杆身的结构形状。基于上述考虑,我们在杆身侧边上选择 6 个控制点作为设计变量来控制杆身的形状。其中  $x_6$  取为固定值,因此实际设计变量数目为 5 个,即  $x_1, x_2, x_3, x_4$  和  $x_5$ ,这样待优化的连杆形状由  $x_1, x_2, x_3$  和  $x_4, x_5, x_6$  组成的二次曲线以及由  $x_3, x_4$  组成的直线来描述。

3) 约束条件 为了保证连杆工作时安全可靠,连杆上节点的最大有效应力必须小于其许用应力值。其有效应力按下式计算:

$$\delta_{ef} = \sqrt{\delta_x^2 + \delta_y^2} - \delta_x \delta_y + 3\tau_{xy}^2 \quad (44.6-44)$$

##### (3) 形状优化模型求解

将不同的敏度分析方法用于连杆的形状优化设计中,优化方法调用优化方法程序库,优化结果见表 44.6-1。

表 44.6-1 连杆形状优化设计中四种敏度分析方法结果比较

敏度分析方法	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	目标函数 $W$ /kg	CPU 运行时间 $t$ /min
差分法	5.150	3.251	3.072	3.320	3.891	3.162	41
拟解析法	5.325	3.231	3.031	3.331	3.845	3.342	24
拟载荷法	4.812	3.292	3.151	3.370	3.980	2.802	20
单位虚载荷法	4.812	3.292	3.151	3.370	3.980	2.802	14

在表 44.6-1 中, 设计变量初值为:  $x_1=5.5, x_2=3.2, x_3=3, x_4=3.3, x_5=3.8$ ; 目标函数初值为  $W=3.513\text{kg}$ 。图 44.6-3 为连杆优化前后的形状, 图 44.6-4 为采用各种不同敏度分析方法的优化迭代过程。从图 44.6-4 中可以看出, 采用单位虚载荷法进行形状优化不但迭代次数少, 而且目标也最优, 计算时间也最少。

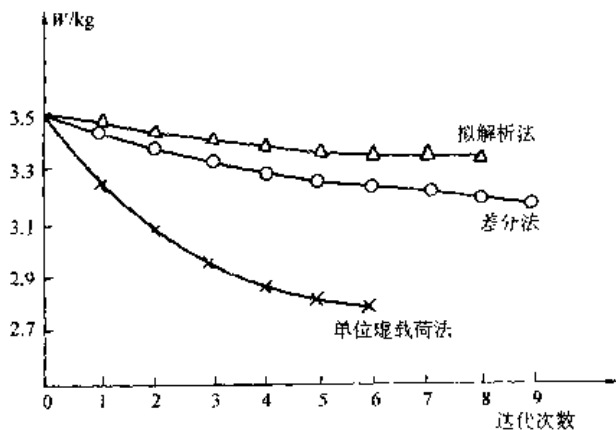


图 44.6-4 连杆优化迭代过程

### 5.2 起重吊钩的自适应形状优化

起重吊钩的几何尺寸、边界条件以及载荷如图 44.6-5 所示; 图 44.6-6 则表示了待优化的形状设计参数, 其优化数学模型为

$$\begin{aligned} \min. \quad & F[X] \approx \min\{\max(\delta_{ef}[X])\} \\ \text{s. t.} \quad & W[X] - W[X_0] \leq 0 \end{aligned} \quad (44.6-45)$$

- 式中  $F[X]$  —— 目标函数;
- $\delta_{ef}[X]$  —— 吊钩体内有效应力;
- $W[X]$  —— 吊钩重量;
- $W[X_0]$  —— 吊钩初始重量。

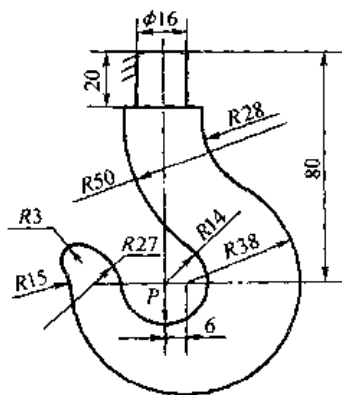


图 44.6-5 吊钩的几何尺寸、载荷及边界条件

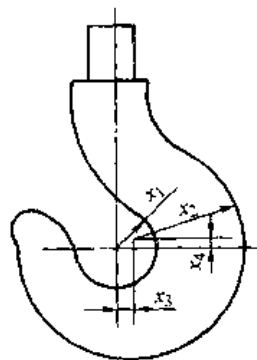


图 44.6-6 吊钩的形状设计变量

对于这一优化问题, 采用了自适应分析的形状优化及不采用自适应分析的形状优化两种方法。

表 44.6-2 给出了这两种情况下设计变量的初值、优化值及设计变量的上、下限。图 44.6-7 表示了这两种优化方法的优化迭代过程。由图 44.6-7 可见, 采用自适应分析后, 目标函数值由  $771.211\text{N/mm}^2$  下降到  $651.189\text{N/mm}^2$ , 下降 15.6%, 并且目标函数在迭代过程中无波动现象, 一直平稳下降, 从优化开始到结束只迭代了 6 次。不采用自适应分析的形状优化, 目标函数值由  $771.211\text{N/mm}^2$  下降到  $680.103\text{N/mm}^2$ , 只下降 11.8%, 而且迭代过程中有波动, 迭代次数也较多。故由此可得出结论, 形状优化中采用自适应分析, 可以使目标函数值下降幅度增大, 优化过程保持良好的稳定性, 并可减少迭代次数。

表 44.6-2 起重吊钩形状优化设计变量的初值及优化值

设计变量	初值	优化值 <sup>①</sup>	优化值 <sup>②</sup>	上限	下限
$x_1$	14	12	12.2	20	12
$x_2$	38	36.3	37.2	50	30
$x_3$	6	10	6.4	10	-10
$x_4$	0	0.8	0.63	10	-10

- ① 采用自适应分析的优化结果;
- ② 不采用自适应分析的优化结果。

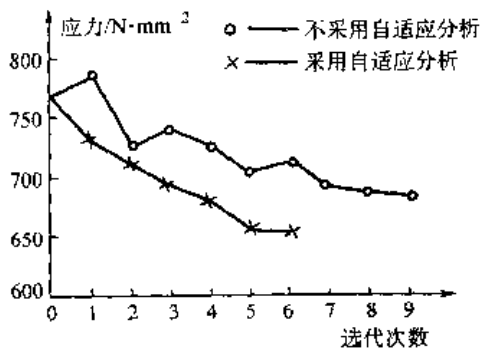


图 44.6-7 吊钩形状优化迭代过程图

# 第 7 章 可靠性优化设计

## 1 可靠性优化设计概述

可靠性是指产品在一定的使用条件下, 在规定的时间内实现确定功能的能力。为了保证产品在一段时间内的工作过程中不发生失效或破坏, 需要采用相应的设计准则和设计方法。机电产品可靠性包括电子产品可靠性、控制系统(包括软件和硬件系统)可靠性和机械产品可靠性(包括结构可靠性和机械可靠性)。利用加大安全系数来改善产品或系统可靠性的手段难以达到应有效果, 必须采取可靠性设计方法进行产品或系统的设计。

可靠性设计是以概率论、数理统计为基础结合专业领域设计方法而形成的一种综合性设计技术。可靠性设计是将常规设计方法中所涉及到的设计变量(如载荷、应力、强度、寿命等)所具有的多值现象看成是服从某种分布规律的随机变量, 用概率统计方法设计出符合产品可靠性指标要求的主要参数和结构尺寸。可靠性优化设计是以可靠性设计和优化技术为基础, 将可靠性理论与优化设计结合起来的设计方法。目前, 可靠性优化设计可以分为两类: 一类是以可靠性指标为优化目标的优化设计; 另一类是以可靠性为约束条件的优化设计问题。

以可靠性指标为优化目标的优化设计模型如下式所示:

$$\begin{aligned} \max. & R(x) \\ \text{s.t.} & \left. \begin{aligned} g_i(x) &= 0 \quad i = 1, 2, \dots, q \\ g_j(x) &\leq 0 \quad j = q + 1, q + 2, \dots, p \\ x^l &\leq x \leq x^u \end{aligned} \right\} \end{aligned} \quad (44.7-1)$$

式中  $R(x)$  —— 零件的可靠度;

$x$  —— 设计变量;

$g_i(x)$  —— 约束条件;

$x^l, x^u$  —— 设计变量  $x$  的下界值和上界值。

以可靠性为约束条件的优化设计模型如下式所示:

$$\begin{aligned} \min. & C(x) \text{ 或 } f(x) \quad x \in R^n \\ \text{s.t.} & \left. \begin{aligned} R(x) &\geq R^0 \\ g_i(x) &= 0 \quad i = 1, 2, \dots, q \\ g_j(x) &\leq 0 \quad j = q + 1, q + 2, \dots, p \\ x^l &\leq x \leq x^u \end{aligned} \right\} \end{aligned} \quad (44.7-2)$$

式中  $C(x)$  —— 设计对象的价格函数;

$f(x)$  —— 设计对象的性能函数;

$R^0$  —— 设计对象的可靠度要求。

可靠性优化设计已经成为优化设计的重要分支。目前在可靠性优化设计方面, 国内学者已经开展了许多研究工作。在常用机械零部件齿轮、减速器方面, 国内已经开展了齿轮传动的可靠性优化设计、行星齿轮传动的可靠性优化设计、直齿锥齿轮传动的可靠性优化设计、圆弧齿圆柱齿轮传动的可靠性优化设计、圆柱齿轮减速器可靠性优化设计、弧齿锥齿轮传动可靠性优化设计、2K- $\Theta$ 行星齿轮减速器的可靠性优化设计、工程机械行星减速器可靠性优化设计、斜齿圆柱齿轮传动的可靠性优化设计、电梯圆柱蜗杆传动的可靠性优化设计、NGW型行星减速器可靠性优化设计、圆柱蜗杆减速器的可靠性优化设计、蜗杆传动的可靠性优化设计、齿轮接触强度的可靠性优化设计、中速磨煤机减速器可靠性优化设计、普通圆柱蜗杆减速器的可靠性优化设计、航空齿轮传动的可靠性优化设计等工作。在弹簧、联接件方面, 已经开展了交变载荷碟形弹簧的可靠性优化设计、汽车离合器圆柱螺旋式压簧的可靠性优化设计、圆柱形扭转弹簧的可靠性优化设计、铁道车辆螺旋支承弹簧的可靠性优化设计、发动机气门弹簧的可靠性优化设计、圆柱形扭转螺旋弹簧的可靠性优化设计、螺旋管簧的可靠性优化设计、电器弹簧的可靠性优化设计、车辆用钢板弹簧的可靠性优化设计、变载荷螺栓组联接的可靠性优化设计、联接螺栓组的可靠性优化设计、楔环联接的可靠性优化设计、圆柱面过盈联接的可靠度优化与可靠性优化设计、汽车发动机连杆螺栓的可靠性优化设计、滚子输送链联接牢固度的可靠性优化设计、压力容器螺栓组联接的可靠性优化设计、船舶轴系法兰联接螺栓组的可靠性优化设计、注塑机合模油缸螺栓组的可靠性优化设计等工作。在专业机械、机械结构等方面, 国内学者开展了装载机差速器的可靠性优化设计、球磨机传动系统的可靠性优化设计、齿带传动的可靠性优化设计、滚子链传动的可靠性优化设计、变速器的可靠性优化设计、工程机械传动轴的可靠性优化设计、火电厂推车驱动装置的可靠性优化设计、矩形截面钢筋混凝土筒支梁的可靠性优化设计、两端任意支承的矩形截面钢筋混凝土梁的可靠性优化设计、中心受压钢筋混凝土矩形截面柱的可靠性优化设计、工型截面框架结构基于可靠

性的优化设计、薄壁梁的可靠性优化设计、有加强圈外压薄壁圆筒的可靠性优化设计、框架结构的可靠性优化、离散变量框架结构的可靠性优化、机械系统可靠性优化设计策略、响应面法的可靠性优化设计策略、大型结构可靠性优化设计的系统方法、工程机械可靠性优化设计、流体动压润滑向心滑动轴承的可靠性优化设计、转轴可靠性优化设计、塔式起重机塔头结构的可靠性优化设计、牙轮钻气控系统的可靠性优化设计、机械产品全寿命期价值期望的可靠性优化模型、钢筋砼框架结构的可靠性优化设计研究、后桥壳的可靠性优化设计、汽车前轴的可靠性优化设计、结构动力响应可靠性优化算法、内燃机连杆可靠性优化设计、扭杆的可靠性优化设计、压力容器的可靠性优化设计、整体法兰可靠性优化设计、空心传动轴的可靠性优化设计等工作。

模糊分析、遗传算法、神经网络等技术也用来解决可靠性优化设计问题，在这些方面已经开展了圆柱齿轮传动模糊可靠性优化设计、基于模糊分析的蜗杆传动的可靠性优化设计、蜗杆传动的多目标模糊可靠性优化设计、基于模糊分析的渐开线少齿差行星传动的可靠性优化设计、摆动活齿传动的双目标模糊可靠性优化设计、模糊带传动的可靠性优化、弹簧模糊可靠性优化设计、基于模糊理论的球磨机传动系统的可靠性优化设计、基于遗传算法的可靠性优化、简易海洋平台疲劳可靠性优化设计的进化算法、基于遗传算法的动力箱可靠性优化设计、基于遗传算法的机械零部件可靠性优化设计、基于模拟退火算法的可靠性优化、遗传算法的复杂系统可靠性优化设计、利用遗传算法和模糊逻辑进行系统可靠性优化分配、遗传算法求解系统可靠性优化、基于神经网络理论的船舶结构可靠性优化设计、基于神经网络的内燃机零件可靠性优化设计、基于神经网络的船舶主动力装置系统可靠性优化分配等工作。

本节介绍可靠性优化设计中常见的机构可靠性优化、结构可靠性优化等问题。

## 2 机构可靠性优化设计

机构是由构件和运动副等组成，构件通过运动副来实现它们之间的可动联接。机构是机械设备和系统的重要组成部分。在机械设备中，采用机构可实现预期的运动和传递动力要求。机构可靠性可以认为是机构在一定使用条件和期限下能精确、及时、协调地完成规定机械动作的能力。与其他部件可靠性设计相比，机构可靠性设计更侧重于机构动作在几何空间中的运动精确度、在时间方面的精确性、机构在几何空间和时间方面协调性和同步性等。另外，机构可靠性还需要考虑机构在一个动作循环周期的精确性、及时性和协调性。

机构可以实现多种功能。通过机构可以实现的运动不仅包括减速、变速、换向等运动状态方面的功能，还可以实现转动、平移、空间复合运动等运动形式间转换方面的功能。另外，通过机构可以实现动力的传递，即将动力源输出的功率和转矩传递到执行构件，使执行机构能克服阻力而做功。因此在机构可靠性设计时，不仅需要考虑强度和刚度可靠性方面的要求，还要考虑机构实现运动的可靠性方面的要求，即机构功能可靠性要求。只有确保机构各构件的可靠性，才有可能保证机构不发生构件断裂等破坏性失效现象。在机构可靠性设计时，机构动作或功能性能方面的可靠性更为重要。因为机构在工作过程中虽然没有发生断裂等失效现象，但会因机构磨损等原因致使功能下降而不能完成预定的功能。例如，在机床过度磨损、机床传动机构和进给机构静态动态误差很大情况下，无法加工技术指标等级合格的工件。因此，机构可靠性设计时应特别注意使性能保持在允许值范围内的要求。

一般来说，影响机构工作可靠性的因素很多，主要影响因素包括机构的工作原理、机构动力源特性（例如驱动元件的特性变化）等。机构可靠性分析涉及机构运动学、机构动力学、摩擦磨损理论、可靠性工程等多学科知识。因此以机构可靠性分析为基础的机构可靠性优化将涉及更多的学科领域。本节以连杆机构运动精度可靠性设计为例，介绍连杆机构可靠性优化设计。

连杆机构是典型的运动装置，在设计、生产和使用过程中随机因素的影响，机构各构件的几何尺寸和铰链间隙方面的误差表现为随机性，因此机构运动精度的可靠性优化设计十分重要。在连杆机构中，如果杆长等几何尺寸误差和铰链间隙误差的随机分散性越小，那么机构运动精度可靠性将会越高。由于机构使用中受成本等实际因素制约，因此要求机构的运动精度在满足给定的可靠性约束条件下应尽可能降低制造费用。本节介绍的连杆机构可靠性优化设计，是以具有杆长尺寸和铰链间隙误差的平面四杆机构为对象，以杆长误差和铰链间隙的均方差为设计变量，以运动精度可靠度和设计变量上下限要求为约束条件，以机构制造加工费用最小为目标函数的优化设计数学模型。在机构可靠性分析中，采用微小位移线性叠加原理对机构的输出运动误差进行分析。

在图 44.7-1 所示的平面四杆机构中，设杆 1 为主动件，当杆 1 与  $x$  轴的夹角为  $\theta_1$  时，杆 2 和杆 3 相对于起始位置所转过的角度分别为  $\theta_2$  和  $\theta_3$ ；以  $R_i$  ( $i=1\sim 4$ ) 代表四连杆所形成的四边形矢量图中各边的矢量。

平面四杆机构在设计、生产过程中受多种随机因素的影响，各杆长尺寸和各铰链间隙不可避免地存在着误差，将这些误差处理为随机变量。考虑到加工工艺

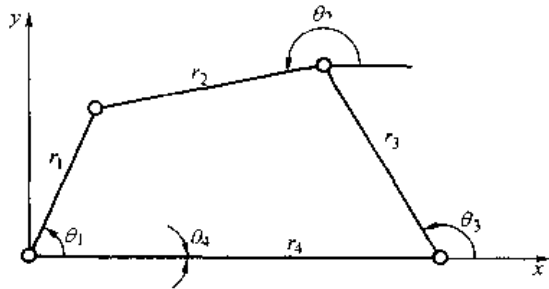


图 44.7-1 平面四杆机构

和方式的不同, 将各杆长尺寸误差随机变量和各铰链间隙误差随机变量处理为相互独立变量, 并与机构几何尺寸相比均为微小量。因此依据微小位移误差的线性叠加原理, 可以将由杆长尺寸误差和由铰链间隙引起的机构运动输出误差分别求解, 然后将两类误差进行叠加求得机构运动输出的总误差随机变量:

$$\theta_{e3} = \theta_{e2} + \theta_{i3} \quad (44.7-3)$$

式中  $\theta_{e2}$ 、 $\theta_{i3}$ ——分别为由所有杆长尺寸误差和铰链间隙而导致的机构运动输出总误差随机变量。

利用矩法和随机变量矩的性质, 可求得由于杆长尺寸误差和铰链间隙而导致机构运动输出的最大总误差  $\theta_{e3}$  的均值  $\mu\theta_{e3}$  和方差  $\sigma^2\theta_{e3}$  分别为

$$\mu\theta_{e3} = \mu\theta_{e2} + \mu\theta_{i3} \quad (44.7-4)$$

$$\sigma^2\theta_{e3} = \sigma^2\theta_{e2} + \sigma^2\theta_{i3} \quad (44.7-5)$$

式中  $\mu\theta_{e2}$ 、 $\sigma^2\theta_{e2}$ 、 $\mu\theta_{i3}$ 、 $\sigma^2\theta_{i3}$ ——分别为由所有杆长尺寸误差和所有铰链间隙而导致机构运动输出误差随机变量的均值和方差。

若给定四杆机构运动输出的许可总误差  $\Psi$  是随机变量, 其均值和方差分别为  $\mu\Psi$  和  $\sigma^2\Psi$ , 则利用可靠性分析的一次二阶矩方法, 机构运动精度的可靠性指标  $\beta$  和相应的可靠度  $P_r$  分别为

$$\left. \begin{aligned} \beta &= (\mu\Psi - \mu\theta_{e3}) / (\sigma^2\Psi + \sigma^2\theta_{e3})^{-1/2} \\ P_r &= \Phi(\beta) \end{aligned} \right\} \quad (44.7-6)$$

式中  $\Phi(\cdot)$ ——表示标准正态分布函数。

平面四杆机构的可靠性优化设计应能实现要求的运动规律, 因此机构中各杆长的公称尺寸实际上已由设计要求的运动规律可以确定。这里选择了各杆长尺寸误差和各铰链间隙误差两者的均方差作为设计变量。这是因为对尺寸和间隙误差均方差的要求实质即为对机械加工精度的要求, 对它们进行优化设计, 就意味着对杆长尺寸和铰链间隙的加工精度进行优化, 从而最终实现了对机构运动精度的优化。

假设各杆长尺寸  $r$  和各铰链间隙  $\delta$  均为服从正态分布的随机变量, 其最大、最小值分别为  $r_{\max}$ 、 $r_{\min}$ 、 $\delta_{\max}$

和  $\delta_{\min}$ , 则各杆长尺寸误差  $\Delta r$  和各铰链间隙  $\Delta\delta$  分别为

$$\left. \begin{aligned} \Delta r &= r_{\max} - r_{\min} \\ \Delta\delta &= \delta_{\max} - \delta_{\min} \end{aligned} \right\} \quad (44.7-7)$$

由正态分布的对称性和“ $3\sigma$ ”准则, 则  $\Delta r$ 、 $\Delta\delta$  的均值  $\mu_r$ 、 $\mu_\delta$  和均方差  $\sigma_r$ 、 $\sigma_\delta$  分别为

$$\left. \begin{aligned} \mu_r &= (r_{\max} + r_{\min}) / 2, R_r = \Delta r / 6 \\ \mu_\delta &= (\delta_{\max} + \delta_{\min}) / 2, R_\delta = \Delta\delta / 6 \end{aligned} \right\} \quad (44.7-8)$$

在优化设计中, 是给定各杆长尺寸和各铰链间隙的均值  $\mu_{r_i}$ 、 $\mu_{\delta_i}$  ( $i=1\sim 4$ ), 而以它们的均方差  $\sigma_{r_i}$ 、 $\sigma_{\delta_i}$  ( $i=1\sim 4$ ) 作为设计变量。

机构可靠性优化设计的目的是尽可能地降低机构的制造加工费用, 因而选择了以制造费用最小化作为目标函数。从机械加工的常识可知, 欲使尺寸和配合误差的均方差(公差)越小, 则加工的精度要求将越高, 相应的加工成本和费用也将随之增加。将  $\Delta r$  和  $\Delta\delta$  从小到大等间隔的划分为若干个精度等级, 按精度等级越高, 制造费用越高的原则构造费用函数, 即将第  $i$  个杆长和第  $j$  个铰链的加工费用函数分别表示为

$$\left. \begin{aligned} G_i &= G_i(\sigma_i) \\ G_j &= G_j(\sigma_j) \end{aligned} \right\} \quad (44.7-9)$$

下面讨论构造  $G_i(\sigma_i)$  和  $G_j(\sigma_j)$  的具体表示形式。不论采用何种机械加工设备, 只要其工作就将有一个基本费用。此外, 随着加工精度  $X_i$  的提高, 其制造加工费用一般均呈非线性规律单调增长。因此可假设某零件的加工精度  $X_i$  与制造加工费用  $G(X_i)$  之间为如下二次函数关系:

$$G(X_i) = C_i(X_i - E_i)^2 + D_i \quad (44.7-10)$$

式中  $C_i$ ——设备的加工等级费用修正系数;

$D_i$ ——设备的基本费用;

$E_i$ ——最低加工精度。

由机械加工工艺可知, 对于相同的加工精度要求构件各自加工工艺难易程度不同, 它们的加工费用是有差异的, 这可通过调整系数  $C_i$  来反映。另外, 不同的设备加工相同精度要求的构件, 其费用函数有大小和变化快慢之分, 对此可通过改变系数  $C_i$  和  $D_i$  来调整费用函数曲线。

分别建立了平面四杆机构的杆长加工精度的费用函数  $G(\sigma_r)$  和铰链加工精度的费用函数  $G(\sigma_\delta)$ , 并将两者综合获得平面四杆机构的加工费用函数, 即优化设计中的目标函数为

$$\begin{aligned} G(\sigma) &= G(\sigma_r) + G(\sigma_\delta) \\ &= \sum_{i=1}^4 [C_{r_i}(\sigma_{r_i} - E_{r_i})^2 + D_{r_i}] \\ &\quad + \sum_{j=1}^4 [C_{\delta_j}(\sigma_{\delta_j} - E_{\delta_j})^2 + D_{\delta_j}] \end{aligned} \quad (44.7-11)$$

式中  $E_i = E_{\theta_i} = 1.0$ , ( $i = 1 \sim 4$ );  $D_i = D_{\theta_i} = 1$ , ( $i = 1 \sim 4$ );  $C_{r1} = 425478.7926$ ;  $C_{r2} = 173707.7601$ , ( $i = 2 \sim 4$ );  $C_{\delta j} = 173707.7601$ , ( $j = 1 \sim 4$ )。

机构可靠性优化设计的约束条件不同于常规的确定性的机构精度优化设计问题。以机构运动精度的可靠性指标大于等于设计给定的可靠性指标及各设计变量(即四杆长尺寸误差的均方差和四铰链间隙误差的均方差)不超出其相应给定的上下限作为约束条件。约束条件可以采用下式描述:

$$\begin{aligned} \min: & \quad G(X) = \sum_{i=1}^4 [C_{r_i}(\sigma_{r_i} - E_{r_i})^2 + D_{r_i}] + \sum_{i=1}^4 [C_{\delta_i}(\sigma_{\delta_i} - E_{\delta_i})^2 + D_{\delta_i}] \\ & \quad (X) = (\sigma_{r_1}, \sigma_{r_2}, \sigma_{r_3}, \sigma_{r_4}, \sigma_{\delta_1}, \sigma_{\delta_2}, \sigma_{\delta_3}, \sigma_{\delta_4}) \\ \text{s. t.} & \quad (\mu\Psi - \mu\theta_{\theta_3})(\sigma^2\Psi - \sigma^2\theta_{\theta_3})^{-1/2} - \beta^* \geq 0 \\ & \quad \sigma^i \leq \sigma_i \leq \sigma^o (i = 1 \sim 8) \end{aligned} \quad (44.7-12)$$

式中  $\beta^*$ ——设计给定的可靠性指标,它由设计给定的可靠度  $P^*$  确定:  $\beta^* = \phi^{-1}(P^*)$   
 $\sigma^o, \sigma^i$ ——分别为设计变量  $\sigma_i$  的上下限,即限定的最高和最低加工精度。

平面四杆机构运动精度可靠性优化设计数学模型可以描述为

在机构可靠性优化数学模型中,以运动精度的可靠性指标作为机构的性态约束,且约束函数又是设计变量的复合函数,因此采用外点法将约束优化问题转化为无约束优化问题,然后利用直接优化方法中的 Powell 方向加速方法求解。

假设各杆长尺寸分别为:  $r_1 = 0.08\text{m}$ ,  $r_2 = 0.24\text{m}$ ,  $r_3 = 0.208\text{m}$ ,  $r_4 = 0.2\text{m}$ ,且各杆长尺寸误差的均值为零,其偏差呈对称性。各铰链间隙的均值均为  $q_{\delta i} = 1.0$

$\times 10^{-4}\text{m}$ ,其偏差也为对称性。设计给定机构输出运动  $\theta_3$  的许可误差为确定量  $\Psi = 0.5^\circ$ ; 给定的机构运动精度可靠度为  $P^* = 0.95053$ ,其对应的可靠性指标为  $\beta^* = 1.65$ 。各杆长精度的允许范围为  $1.333 \times 10^{-7}\text{m} \leq \sigma_{r_i} \leq 7.667 \times 10^{-4}\text{m}$ ,  $3.3 \times 10^{-7}\text{m} \leq \sigma_{r_i} \leq 1.2 \times 10^{-1}\text{m}$  ( $i = 2 \sim 4$ )。各铰链间隙精度的允许范围为  $1.667 \times 10^{-7}\text{m} \leq \sigma_{\delta_i} \leq 1.667 \times 10^{-5}\text{m}$  ( $i = 1 \sim 4$ )。选择两个设计解为初始点,优化结果如表 44.7-1 所示。

表 44.7-1 平面四杆机构精度可靠性优化设计结果 ( $\beta^* = 1.65$ )

设计变量	初始设计 I	优化结果	初始设计 II	优化结果
$\sigma_{r1}$	$7.66667 \times 10^{-4}$	$6.69832 \times 10^{-4}$	$1.3333 \times 10^{-7}$	$6.69759 \times 10^{-4}$
$\sigma_{r2}$	$1.2 \times 10^{-3}$	$6.71769 \times 10^{-3}$	$3.3333 \times 10^{-7}$	$6.71752 \times 10^{-4}$
$\sigma_{r3}$	$1.2 \times 10^{-3}$	$1.13585 \times 10^{-3}$	$3.3333 \times 10^{-7}$	$1.13589 \times 10^{-3}$
$\sigma_{r4}$	$1.2 \times 10^{-3}$	$9.2369 \times 10^{-4}$	$3.3333 \times 10^{-7}$	$9.23773 \times 10^{-4}$
$\sigma_{\delta 1}$	$1.66667 \times 10^{-5}$	$7.45534 \times 10^{-6}$	$1.66667 \times 10^{-7}$	$7.43293 \times 10^{-6}$
$\sigma_{\delta 2}$	$1.66667 \times 10^{-5}$	$7.44352 \times 10^{-6}$	$1.66667 \times 10^{-7}$	$7.42941 \times 10^{-6}$
$\sigma_{\delta 3}$	$1.66667 \times 10^{-5}$	$7.43383 \times 10^{-6}$	$1.66667 \times 10^{-7}$	$7.42944 \times 10^{-6}$
$\sigma_{\delta 4}$	$1.66667 \times 10^{-5}$	$7.45534 \times 10^{-6}$	$1.66667 \times 10^{-7}$	$7.43293 \times 10^{-6}$
目标函数 G	8	10.3938	44.0068	10.3938
可靠性指标 $\beta$	1.11089	1.65	2886.79	1.65

### 3 结构可靠性优化设计

结构是指若干元件组成的能承受外部作用并有特定功能的整体。在结构中,各个元件之间存在相互作用和相互依存的关系。利用结构可靠性数量指标就可以对结构进行可靠性评价。一般来说,结构系统可靠性随

着使用时间的增加而降低。

由于结构中各个元件之间存在相互关联关系,结构可靠性分析、结构设计就显得十分困难。一般在进行结构可靠性分析计算时,先确定结构的失效模式和主要失效模式。结构系统的种类很多,因此结构的失效模式也各不相同。对于高度静不定结构,一般多个元件达



到临界状态或破坏后整个结构才会失效，这就是一个失效模式。对于大型复杂结构系统，失效模式数量多，例如十六杆平面超静定结构最大可能的失效模式就有上万种，因此需要确定结构的主要失效模式。所谓结构的主要失效模式，就是与一般失效模式相比该模式发生失效的概率比单个失效模式的失效概率高两到三个数量级。与主要失效模式相比，一般失效模式对结构的影响在可靠性计算过程中可以不予考虑。

结构可靠性优化设计是以结构可靠性分析和结构优化设计为基础发展起来的设计方法。采用结构可靠性优化设计，可以得到比较均匀的安全水平或可靠性水平的结构设计。以结构可靠性分析为基础进行优化设计，可以将可靠度设计的要求反映到优化设计问题的约束函数或目标函数中，从而得到更能反映实际使用状况的结构最佳设计方案。

在结构可靠性优化设计中优化模型十分关键。结构可靠性优化模型包括结构可靠性分析模型和优化模型；利用结构可靠性分析模型可以进行结构响应计算和结构可靠性估计，并为优化计算提供结构响应参数、状态约束以及对设计变量的梯度等；利用优化模型可以进行包括优化变量、目标函数及约束条件有关可靠性指标的描述。在结构可靠性优化设计中，一般以结构的重量、结构的各种成本、结构的失效损失费用等作为优化的目标函数，以与结构可能失效模式相关联的可靠性要求为约束条件（例如位移、应力、稳定性等方面设计要求）来建立模型进行优化求解。

结构可靠性优化设计模型一般可用下式描述：

$$\begin{aligned} \min. & f(X) \\ \text{s. t.} & \\ & g_j(X) = P_j^* - P\{Z_j(X) \geq \alpha_j\} \leq 0, \\ & \left. \begin{aligned} & h_k(X) \leq 0, \\ & x_{\min} \leq x_i \leq x_{\max} \end{aligned} \right\} \begin{aligned} & j = 1, 2, \dots, m \\ & k = 1, 2, \dots, q \\ & i = 1, 2, \dots, n \end{aligned} \end{aligned} \quad (44.7-14)$$

式中  $f(X)$ ——目标函数， $X$  为结构可靠性优化设计变量；

$g_j(X)$ ——与结构可靠性有关的约束函数；

$h_k(X)$ ——与结构有关的一般性约束函数；

$P_j^*$ ——第  $j$  个可靠性约束中给定的可靠度；

$P\{Z_j(X)\}$ ——结构的实际可靠度；

$Z_j(X)$ ——第  $j$  个可靠性约束对应的功能函数（随机变量概率分布）；

$\alpha_j$ ——给定的功能函数  $Z_j(X)$  界限值。

在结构可靠性优化设计模型中，位移、应力、稳定性等结构方面的约束函数可以处理为概率形式约束（即  $g_j(X)$ ），也可以为非概率形式的约束（即  $h_k(X)$ ）。

为了进行可靠性计算方便起见，对于载荷、强度等随机参数的分布可以按照正态分布来处理，对于作用于结构系统的载荷和结构元件强度间的关系，也可以认为是相互独立的。结构的失效概率等计算，一般可采取数值积分方法、一阶或二阶可靠性方法、Monte Carlo (MC) 法等。采用 Monte Carlo 法，可按照基本随机变量的概率密度函数抽取样本，并代入功能函数进行统计试验而求解。对大规模的数值仿真，Monte Carlo 法计算量很大，并不实用。在结构可靠性优化中，需多次重复估算结构体系的可靠性，并进行敏度计算。结构体系的可靠性分析，一般涉及结构主要失效模式、单失效模式的可靠性计算、结构体系的可靠性估算。结构体系的可靠性主要由发生概率相对很高的少量主要失效模式决定，所以主要失效模式是结构体系可靠性分析中的关键问题之一。比较常用的方法有  $\beta$  展开法、分枝限界法、工程准则和优化准则法等。分枝限界法是通过分枝找出全部失效模式的失效路径，再通过限界处理舍弃其中对体系可靠性贡献很小的次要失效模式，仅保留失效模式所对应的失效路径。该方法对冗余度较低的结构系统很有效，是中、小规模结构可靠性优化的常用方法之一。准则法是用增量载荷法确定体系的强度，用元件的承力比或承力比的变化率确定主要失效模式。由于优化准则法的计算量小、效率高，也是可靠性优化的常用算法之一。

关于结构可靠性优化设计方法，一般可选择非线性规划的优化算法。与一般结构优化设计问题相似，各种优化方法的效率和有效性对可靠性优化问题而言相差很大。直接处理约束的优化方法如罚函数法、序列线性规划法、序列二次规划法、梯度法、拉格朗日法及可行方向法等，对小规模的结构可靠性优化问题比较适用。可靠性计算及优化过程中计算量十分大，因此近似计算策略对结构可靠性优化计算很重要。一些有效的近似方法被引入结构可靠性优化设计领域。这些近似方法通过结构响应、功能函数的构造及可靠性计算等方面的近似和简化处理，使原来复杂的计算过程可用于复杂结构的优化。

下面简单介绍图 44.7-2 所示三杆结构可靠性优化设计实例。

图 44.7-2 所示意的三杆结构是超静定结构，假设材料强度的屈服点  $C_m$  和结构载荷  $L_1$  和  $L_2$  为相互独立的正态随机变量，各杆长度为确定量  $l_q$  ( $q=1, 2, 3$ )，各杆截面积为设计变量  $A_q$  ( $q=1, 2, 3$ )。下面进行三杆结构材料费用最小的可靠性优化设计。

先考虑结构的失效模式。如图 44.7-2 所示结构有三种失效模式：

模式 1：1、2 杆失效

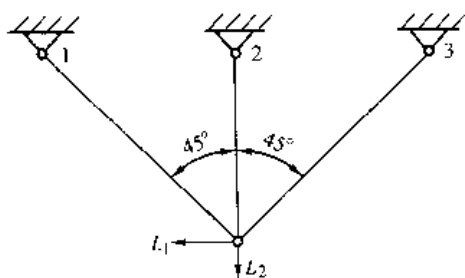


图 44.7-2 三杆结构

$$Z_1 = R_1 + \frac{\sqrt{2}}{2}R_2 + R_3 - \frac{\sqrt{2}}{2}L_1 - \frac{\sqrt{2}}{2}L_2 \quad (44.7-15)$$

模式 2: 2、3 杆失效

$$Z_2 = \frac{\sqrt{2}}{2}R_1 + R_2 - \frac{\sqrt{2}}{2}L_1 - \frac{\sqrt{2}}{2}L_2 \quad (44.7-16)$$

模式 3: 1、3 杆失效

$$Z_3 = \frac{\sqrt{2}}{2}R_1 + \frac{\sqrt{2}}{2}R_2 - L_1 \quad (44.7-17)$$

各杆的载荷为  $R_q = C_q A_q$  ( $q=1, 2, 3$ )，表 44.7-2 给出了  $R_q$  的变异系数  $V$  的值，失效概率  $P$  可由有关公式计算。

设结构的初始成本仅与结构的质量有关，则其初

始成本可写为

$$C_0(A_q) = \sum_{q=1}^3 \mu_q \rho_q l_q A_q \quad (44.7-18)$$

式中  $\mu_q$  第  $q$  个杆件每单位质量材料的成本；

$\rho_q$  第  $q$  个杆件的材料密度；

$l_q$  第  $q$  个杆件的长度；

$A_q$  第  $q$  个杆件的横截面积。

假设期望的结构失效成本为  $C_f(P_1) = C_1 P_1$ ， $C_1$  为估算的结构失效后的总损失，因而能够得出期望的结构总成本为

$$C(A_q) = \sum_{q=1}^3 \mu_q \rho_q l_q A_q + C_1 P_1(A_q) \quad (44.7-19)$$

用优化方法可计算期望结构总成本最小的设计变量的最优解， $C_1$  取各种值时优化结果如表 44.7-3 所示。

表 44.7-2 三杆结构数据

$q$	$l_q/\text{cm}$	$V_{R_q}$	$C_{x_q}/\text{MPa}$
1	215.5	0.05	275.8
2	152.4	0.05	275.8
3	215.5	0.05	275.8

表 44.7-3 优化结果

$C_1/\text{元}$	$A_1/\text{cm}^2$	$A_2/\text{cm}^2$	$A_3/\text{cm}^2$	$P_1$	$C_0/\text{元}$	$C/\text{元}$
$10^2$	13.3	21.3	10.1	$9.64 \times 10^{-3}$	14.71	15.67
$10^3$	14.4	22.6	11.4	$7.79 \times 10^{-4}$	16.85	17.63
$10^4$	17.0	23.7	12.4	$6.79 \times 10^{-5}$	18.61	19.29
$10^5$	19.2	24.6	13.3	$6.16 \times 10^{-6}$	20.15	20.77
$10^6$	21.3	25.5	14.1	$5.71 \times 10^{-7}$	21.55	22.13

## 第 8 章 复杂系统优化设计

### 1 复杂系统优化设计概述

复杂系统可以认为是由许多较小的系统（或称为子系统）组成的一个整体。一般认为复杂系统涉及了不同学科领域的问题，例如属于复杂系统的飞机是由机翼、机身、尾翼、推进、电子通信等分系统组成。涉及机械结构、空气动力学、发动机、电子通信等许多技术领域。

复杂系统的设计更具有挑战性，因此更需要采用优化设计等先进的技术手段。一般复杂系统优化设计可以描述如下式所示的优化问题：

$$\left. \begin{array}{l} \min. F(X, U, Y) \\ \text{s. t.} \\ H(X, U, Y) = 0 \end{array} \right\} \quad (44.8-1)$$

式中  $X$ ——复杂系统有关设计的向量；

$U$ ——复杂系统有关状态向量；

$Y$ ——子系统之间相互作用的关联向量。

在复杂系统优化设计问题中， $F(X, U, Y)$  描述了要优化的设计目标， $H(X, U, Y) = 0$  为复杂系统需要满足的约束条件。复杂系统问题往往涉及到多个技术学科领域，因此复杂系统设计优化问题大多属于多学科优化问题，因此可以从多学科设计的角度来处理并描述复杂系统优化设计问题。在式 (44.8-1) 描述的复杂系统优化设计问题中，将控制向量和部分关联向量取为设计变量，一些参量用固定参数描述，子系统之间的部分关联向量的相互耦合采用不等式约束函数  $g_j(X)$  描述，假设复杂系统是由二个子系统  $A$ 、 $B$  组成，则复杂系统优化问题可描述为

$$\begin{array}{l} \min. f = f(X, U, P) \\ \text{s. t.} \\ h_i(X, U, P) = 0 (i = 1, 2, \dots, m) \\ g_j(X, U, P) \leq 0 (j = 1, 2, \dots, n) \end{array}$$

子系统 A

$$\left. \begin{array}{l} \min. f = f(X_A) \\ \text{s. t.} \\ h_{iA}(X_A, U_A, X_B^0, U_B^0) = 0 \\ \quad (i_A = 1, 2, \dots, k_1) \\ g_{jA}(X_A, U_A, X_B^0, U_B^0) \leq 0 \\ \quad (j_A = 1, 2, \dots, k_2) \end{array} \right\} \quad (44.8-2)$$

子系统 B

$$\begin{array}{l} \min. f = f(X_B) \\ \text{s. t.} \\ h_{iB}(X_B, U_B, X_A^0, U_A^0) = 0 \\ \quad (i_B = 1, 2, \dots, k_3) \\ g_{jB}(X_B, U_B, X_A^0, U_A^0) \leq 0 \\ \quad (j_B = 1, 2, \dots, k_4) \end{array}$$

上式中， $f$  表示复杂系统优化设计的目标函数； $X$  表示设计变量向量，设计变量向量由分别与二个子系统  $A$ 、 $B$  相关的设计变量子向量  $X_A$ 、 $X_B$  组成，即  $X = (X_A, X_B)$ ；设计变量子向量间互不重叠； $U$  表示状态向量，状态向量由分别描述子系统  $A$ 、 $B$  性能或特征状态变量子向量  $U_A$  和  $U_B$  组成； $P$  是固定参数； $h_i(X, U, P)$  是等式约束； $g_j(X, U, P)$  是不等式约束。对于子系统  $A$ ， $X_B$ 、 $U_B$  是给定参数，约束条件只包括与子系统  $A$  设计变量  $X_A$  和状态变量  $U_A$  有关的约束，并不是整个系统优化问题中的全部约束。对于子系统  $B$ ， $X_A$ 、 $U_A$  亦是给定参数。

在复杂系统优化设计问题中，状态变量是用来表达系统性能或特征参数。状态变量是影响设计过程的重要参数，状态变量的计算涉及多学科领域设计问题，一般需涉及各技术学科领域的分析模型。复杂系统优化问题中约束函数  $h_j(X)$ 、 $g_j(X)$  及目标函数  $f(X)$  通常是设计变量和状态变量的函数。

对于式 (44.8-2) 描述的复杂系统优化问题，可以先求出了系统  $A$ 、 $B$  优化问题的解；子系统  $A$ 、 $B$  优化问题的解为所有子系统提供新的设计点  $\{X_A^1, X_B^1\}$ ；各子系统新的设计点可组成整个系统的新设计点  $X^1 = (X_A^1, X_B^1)$ ；以新设计点为基础，计算新的状态变量  $(U_A^1, U_B^1)$ ；以新设计点和新状态变量为基础进行各子系统优化问题的求解；进行上述迭代过程直到得到复杂系统优化解为止。

对于复杂系统优化问题，一般可采用已有的优化方法来求解。但在实际问题的求解过程中常常会存在计算效率低等困难，因此国内外优化设计领域的学者一直没有停止探讨新的复杂优化问题求解策略和算法。因此，本章在第二节专门介绍蚁群算法、免疫算法、智能体等，第三节介绍复杂系统优化问题的分级求解策略多级优化技术，本章第四节介绍适合于复杂系统优化设计问题的多学科优化方法。

## 2 若干优化新算法

### 2.1 粒子群优化算法 (PSO)

粒子群优化算法 (Particle Swarm Optimization Algorithm) 是由 Kennedy 和 Eberhart 提出的对鸟群捕食行为进行抽象的优化算法。粒子群优化算法是通用的启发式搜索方法,其本质上是一种由简单个体组成的群落与环境以及个体之间的互动行为的优化算法。粒子群的迁移过程是有方向性的,搜索过程中运用到反馈原理并可以采用并行计算技术,因此具有较高的优化计算效率。

粒子群优化算法主要借助鸟群捕食行为的思想。当一群鸟在某区域随机搜索食物,所有鸟知道自己当前位置离食物多远,那么搜索的最简单有效的策略就是搜寻目前离食物最近的鸟的周围区域。PSO 算法是将这种思想应用于优化问题而形成的。PSO 算法中,每个优化问题的解都是粒子在搜索空间中的位置,粒子还有一个速度值决定它们飞翔的方向和距离,然后粒子群就追随当前的最优粒子在解空间中搜索。

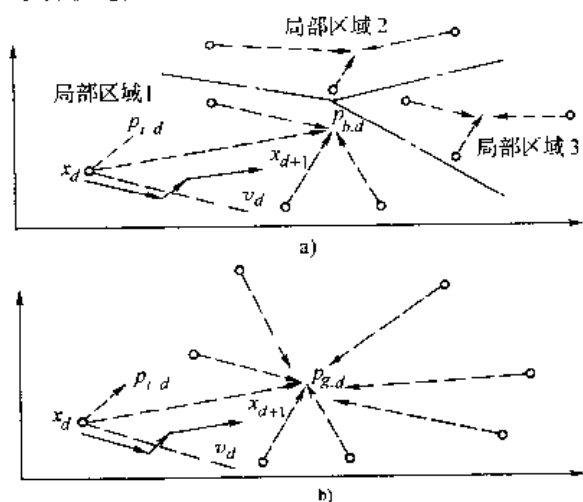


图 44.8-1 粒子群空间寻优过程

a) 粒子群的初始位置和速度值

b) 寻优过程中的粒子群位置和速度

粒子群寻优过程如图 44.8-1 所示。粒子群的初始位置和速度值是随机分布的,如图 44.8-1a 所示,粒子群按空间划分为几个区域。在每次迭代过程中,每个粒子通过跟踪两个“极值”来更新自己在解空间中的位置与速度。第一个极值就是粒子本身在迭代过程中找到的自身的最优解,这个解所对应的粒子叫做个体极值粒子  $p_{i,d}$ ; 另一个极值是各区域局部粒子群中存在的一个最优解,这个解所对应的粒子叫做局部极值粒子  $p_{l,d}$ ,这一寻优过程称为局部 PSO 算法。采用寻优方法是为了避免搜索过程中算法过早地陷入局部最优解。

经过一定的迭代次数后,寻优过程改用全局 PSO 算法加速收敛,如图 44.8-1 所示,整个粒子群通过跟踪个体极值粒子和一个全局极值粒子  $p_{g,d}$  更新自身的速度和位置,在解空间中搜寻全局最优解。粒子寻优遵循以下基本公式:

$$v_{d+1} = c_1 v_d + c_2 (p_{i,d} - x_d) + c_3 (p_{g,d} - x_d) \quad (44.8-3)$$

$$x_{d+1} = x_d + v_{d+1} \quad (44.8-4)$$

式中 下标  $d$ ——迭代次数;

$x_d$ —— $d$  次迭代时的粒子空间位置;

$v_d$ —— $d$  次迭代时的粒子速度;

$p_{i,d}$ ——个体极值粒子;

$p_d$ ——局部极值粒子  $p_{l,d}$  或者全局极值粒子  $p_{g,d}$ ;

$c_1$ 、 $c_2$  和  $c_3$ ——算法的控制参数。

PSO 算法中,每个粒子根据两个极值给出的信息进行变异,这是单向的信息流动,具有方向性,而且 PSO 算法采用实数编码并且没有许多参数需要调整,因此算法的实现更容易。为了避免算法过早陷入局部最优解,在完成迭代寻优操作后引入平移算子规则等启发式变异过程加以解决。PSO 算法流程如图 44.8-2 所示。

### 2.2 智能体 (Agent)

所谓智能体 (Agent),就是指通过传感器感知其周围环境,并通过其行为对环境产生影响的任何实体。Agent 的研究源于人工智能中分布式人工智能,其模型最初是作为一种分布式智能的计算模型提出的。Agent 可以分为智能 Agent、多 Agent 系统 (Multi-Agent System, MAS)、面向 Agent 的程序设计和移动 Agent (Mobile Agent) 四个方面。

尽管目前对 Agent 的定义存在分歧,来自不同领域的研究者从不同的角度对 Agent 做出自己的解释,但人们通常认为一个 Agent 应该具备自治性、反应性、目标性、交互性、主动性、响应性等特性。自治性是指一个 Agent 能在没有与环境的相互作用或来自环境的命令的情况下自主执行任务。主体运行可以不直接由人或者其他东西控制,它对它们自己的行为 and 内部状态有一定的控制权,这是 Agent 区别于普通对象的基本属性。响应性是指 Agent 必须对来自环境的影响和信息做出适当的响应,即对环境的感知和影响。Agent 不仅对环境变化做出反应,而且在特定情况下采取主动行动,这种自身采取主动的能力需要 Agent 有严格定义的目标,表现出面向目标的行为。Agent 的智能由三个主要部件来完成,即内部知识库、学习或自适应能力以及基于知识库内容的推理能力。Agent 可

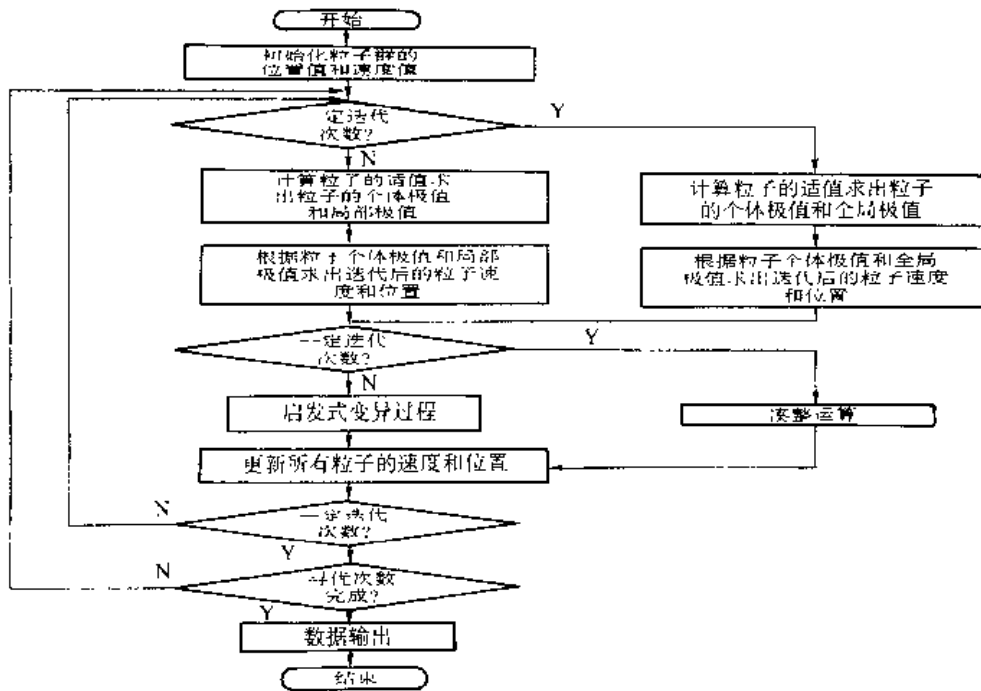


图 44.8-2 算法流程图

以从一个地方移动到另一个地方而保持其内部状态不变, Agent 可以携带数据和能够在远处执行智能指令。Agent 的结构如图 44.8.3 所示。

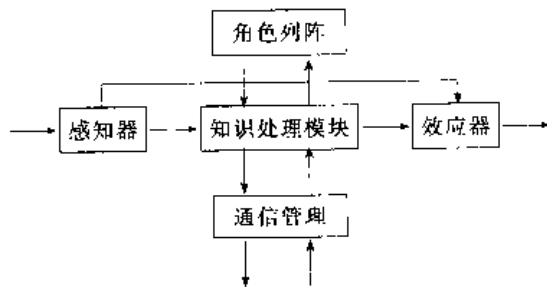


图 44.8-3 Agent 结构

在图 44.8-3 中,感知器是 Agent 接受外部世界信息的感觉通道;知识处理模块是 Agent 的知识库管理系统;通信管理是 Agent 与系统中其他成员进行交互的机制或协议;效应器 Agent 影响或改变环境的界面;角色阵列是 Agent 在 Agent 系统中所扮演角色的列表。

Agent 的结构模型一般包括经典模型、反应模型和混合模型。经典模型是一种形式符号; Wooldridge 和 Jennings 将经典模型定义为“包括显示表示的符号模型,并且其决策是通过逻辑推理、基于模式匹配和符号操作”。经典模型基本上由两部分组成,即规划器和世界模型。经典模型有一个基本的假定,即可以单独研究不同的认知功能(如感知、学习、规划和工作),然后把它们组装在一起构成智能自治 Agent。反应模型是一种基于刺激——行为的反应模型,它是不包含符

号表示的模型,并且不使用复杂符号推理的 Agent。第一个反应模型 Agent 出现在 20 世纪 80 年代中期。反应模型 Agent 思想来源于我们的日常生活。由日常动作而不是抽象推理组成。在这类模型中,仅有一些简单的行为模式,它只是简单地对外部的改变作出反应,不使用复杂的符号推理。反映结构的设计是来自如下假设: Agent 行为的复杂性可以是 Agent 运行环境复杂性的反应,而不是 Agent 复杂内部设计的反应。混合模型结合了前二个模型的优点,可以用来融合经典和非经典的人工智能模型。

MAS 是由一组具有一定资源和能力,相对独立且相互作用的 Agent 组成的计算系统,这些自主的 Agent 通过协作完成某些任务或达到某些目标。MAS 侧重研究如何协调一组 Agent 的行为,即研究这组 Agent 为了在采取行动或求解问题时,如何协调或联合各自的知识、目标、策略和计划等。组成 MAS 的基本单位是 Agent, MAS 中的每个 Agent 都可以实现某种或某些特定的功能,它可以由人控制,也可以代理人的行为在 MAS 中自主地完成电子商品或物理商品的交易过程。图 44.8-4 就是一个包括 3 个 Agent 的简单 MAS。

MAS 的体系结构是指多 Agent 系统中 Agent 之间的信息关系和控制关系、问题求解能力的分布模式。MAS 的体系结构是结构和控制的有机结合,描述了组成 Agent 的基本成分及其作用、各成分的联系及交互机制等。MAS 的体系结构一般也可分为三类,即审慎

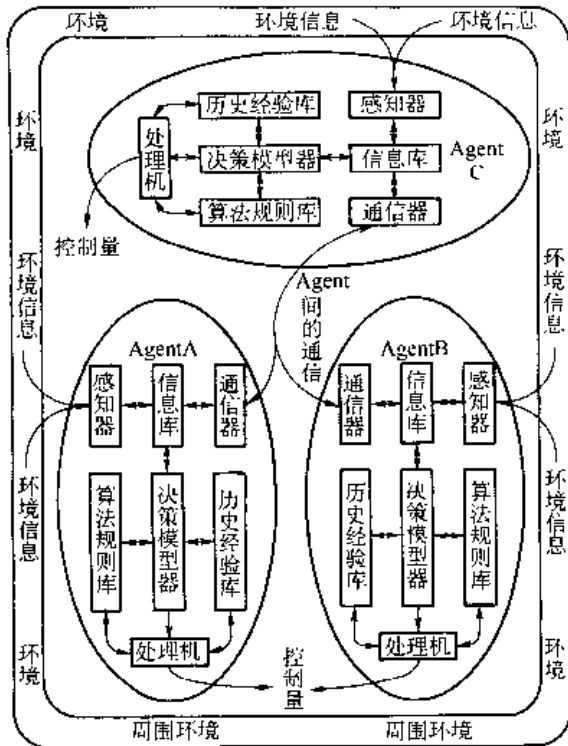


图 14.8-4 MAS 结构框图

式体系结构、反应式体系结构和混合式体系结构。审慎式体系结构包含了显示表示的世界符号模型，Agent 的决策是通过基于模板匹配和符号操作的逻辑推理作出的，如同人们通过深思熟虑、精心推敲作出决策一样，因此称为审慎式体系结构。反应式体系结构包含了感知内外部状态变化的感知器，一组对相关事件作出反应的过程，和一个依据感知器信息激活某过程执行的控制系统。Agent 的活动是由于受到某种刺激而引发的，因此被称为反应式体系结构。该体系结构在目前主流分布式系统中占主导地位。混合式体系结构包含了审慎式和反应式两种结构框架，而且是分层组织的，前者建立在后者的基础上。

工程实践中有许多复杂的优化问题，例如复杂交通网络中车辆行走路线如何；综合考虑人员、设备、产品类型、数量、成本、交货时间等各种因素的生产调度优化；由多个子系统或不同层次系统组成的复杂系统的优化；由多个具有一定智能的智能体组成的系统（即 Multi-Agent System-MAS，如卫星网络、通信网络，空中飞行器、机器人足球赛等）的优化等。这些问题采用一般的优化方法不可能解决，而基于 Agent 的方法对于解决这类复杂问题的优化将显示出其优越性。

### 2.3 免疫算法

免疫算法是基于人工免疫的基本理论发展而来，

它是对人工免疫理论应用和研究的扩充和发展。而人工免疫又是以生物免疫系统基本概念和理论为基础，因此，生物免疫系统的理论是免疫算法的直接来源。免疫算法主要借鉴了免疫系统具有的抗原识别、免疫记忆、免疫调节等特性，将其概念与理论应用于遗传算法中。其中，抗原识别是指通过表达在抗原表面的表位和抗体表面的对位进行相互匹配选择完成识别的过程；免疫记忆是指免疫系统将能与抗原发生反应的抗体作为记忆细胞保存记忆下来，当同类抗原再次入侵时，迅速识别与其发生反应的过程；免疫调节是指在免疫反应过程中，大量的抗体的产生降低了抗原对免疫细胞的刺激，从而抑制抗体的分化和增殖，使抗体和抗原及抗体和抗体之间维持一定强度的免疫平衡。此外，免疫算法的核心内容是免疫疫苗，它是来源于生物疫苗的概念和理论。生物疫苗是医学家利用免疫系统具有的免疫记忆特性发明的，通过注射疫苗，达到快速识别抗原的目的。

免疫算法是在抽取和反映人工免疫理论的基础上，结合遗传算法提出一个用于求解科学研究和工程技术中各种组合搜索和优化计算问题的计算模型。在使用此模型时，将待求解的问题抽象为抗原，待求解问题的解抽象为抗体，而模型中的疫苗则对应于待求解问题解的一些特征信息。

免疫算法核心思想是在合理提取疫苗的基础上，通过接种疫苗和免疫选择两个操作步骤来提高群体适应度，加速迭代过程，并防止群体的退化。

免疫算法：

- 1) 随机产生初始父代种群  $P_0$ ；
- 2) 根据经验知识抽取疫苗；
- 3) 若在当前群体中包含最佳个体，则算法停止运行并输出结果，否则继续；
- 4) 对于目前的第  $k$  代父本种群  $P_k$  进行交叉操作，得到子代种群  $B_k$ ；
- 5) 对  $B_k$  进行变异操作，得到种群  $C_k$ ；
- 6) 对  $C_k$  进行接种疫苗操作，得到种群  $D_k$ ；
- 7) 对  $D_k$  进行免疫选择操作，得到新一代父本  $P_{k+1}$ ，转至 3)。

免疫算法不仅继承了遗传算法的遗传算子，使得免疫算法具有更强的全局搜索能力，而且增加了免疫算子，使得免疫算法有效的防止了遗传算法后期的退化现象，并加快了收敛速度。另外，根据相关定理可以证明免疫算法是概率 1 收敛的，也就是说，免疫算法是强收敛的。

### 2.4 蚁群算法

蚁群算法 (Ant Colony Algorithms) 是由

M1Dorigo 等人在 1991 年提出的优化算法。蚂蚁是相对弱小、功能不强的个体,但是蚂蚁能完成复杂的活动,例如蚂蚁可以寻找到搬运食物的最佳路径。因此一些学者对蚂蚁群体的行为进行了探索,并在此基础上提出了蚁群算法。

蚁群算法是对自然界生物的行为进行抽象而产生的优化算法。以“外激素”(Stigmergy)为媒介间接的联系方式是蚁群中的蚂蚁行为的主要特点,蚂蚁在寻找食物或者寻找回巢的行动中,会在经过的地方留下一些“外激素”的化学物质能被同一蚁群中其他蚂蚁所识别。通过蚂蚁的识别而影响其他蚂蚁的行动,例如蚂蚁会选择留有这些物质的路径进行行动。后来经过的蚂蚁又会留下“外激素”的物质,因此蚂蚁经过的路径上残留了更多的“外激素”。这样蚂蚁经过越多的路径,就会因为残留的外激素浓度较大的缘故被以后蚂蚁选择的可能性就越大。由于越短的路径会被越多的蚂蚁访问,这会一直影响所有蚂蚁都选择最短的路径。自然界蚁群的这种寻找路径的过程表现为一种正反馈的过程,蚁群算法借鉴了上述思路。因此,将基本工作单位视为“蚂蚁”,那么上述寻找路径的过程可以用来进行寻优。另外蚁群算法还具有一定的记忆能力,能够记忆已经访问过的节点;蚁群算法在选择下一条路径的时候并不是完全盲目的,而是按一定的算法有意识地寻找最短路径。

### 3 多级优化设计技术

多级优化设计技术(Multilevel optimization technology)就是将一个复杂的优化设计问题分解成一系列具有多个层次的相对简单的优化问题进行求解的策略。利用多级优化设计策略,可以使许多十分复杂的难以求解的优化设计问题转化为可求解的优化问题。另外,多级优化设计方法与实际设计过程中工程设计人员思维方式是完全相同的。在实际设计过程中,工程设计人员一般先进行总体设计,然后再进行详细的零部件设计。在每个设计阶段都采用优化的思想,即可以形成多级优化设计方法。

多级优化早在 20 世纪 70 年代就曾出现,并被视为解决复杂结构优化问题的一种有效途径。但是通常的多级结构优化都是针对某些特定问题进行的。结构优化中的多级分解主要可分为两大类:一类是基于子结构的分解;另一类是更为采用的基于设计变量层次的分解。多级分解方法,可以有效地对付不同层次的设计变量、目标函数及约束函数。任何一个复杂的结构优化问题,都具有可分层次的潜在属性。

大多数复杂优化问题的目标函数和约束可分离成如下式所示的形式:

$$\left. \begin{aligned} \min_{Y, X} \quad & F(Y, X) = f_0(Y) + \sum_{i=1}^l f_i(Y, x_i) \\ \text{s. t.} \quad & G_{ij}(Y, x_i) \leq 0 \quad i = 1, 2, \dots, l \\ & G_j(Y) \leq 0 \quad j = 1, 2, \dots, J \\ & l = 1, 2, \dots, L \end{aligned} \right\} \quad (44.8-5)$$

式中  $f$  —— 总体目标函数;

$f_0$  —— 与  $Y$  有关的任意函数;

$f_i$  —— 子问题  $i$  的目标函数;

$G_{ij}$  —— 与第  $i$  个子问题相关的约束;

$G_j$  —— 与系统级相关的约束;

$Y$  —— 系统级设计变量;

$x_i$  —— 子问题  $i$  的设计变量;

$J_i$  —— 与第  $i$  个子问题相关的约束个数。

上述问题可以通过两级优化问题来求解,对给定的  $Y$  (由系统级优化给出)求以  $x_i$  为变量的子系统级问题。子系统级优化模型为

$$\left. \begin{aligned} \min_{x_i} \quad & f(Y, x_i) \\ \text{s. t.} \quad & G_{ij}(Y, x_i) \leq 0 \quad i = 1, 2, \dots, l \\ & G_j(Y) \leq 0 \quad j = 1, 2, \dots, J_i \end{aligned} \right\} \quad (44.8-6)$$

然后求解以  $Y$  为设计变量的系统级问题,其中  $x_i$  固定不变(由于已系统级优化求出):

$$\left. \begin{aligned} \min_Y \quad & f(Y, X) = f_0(Y) + \sum_{i=1}^l f_i(Y, x_i) \\ \text{s. t.} \quad & G_j(Y) \leq 0 \quad j = 1, 2, \dots, J \end{aligned} \right\} \quad (44.8-7)$$

通过反复迭代可得  $f(Y, X)$  的最优解。在复杂的结构优化问题中,对不同的结构类型,采用不同的优化方法将会更加有效。优化模型中需要处理目标、约束及设计变量,这三者均与组成结构的单元密切相关,因此可以进行结构优化中的优化单元基类设计等策略。

对优化问题(44.8-5),可以采用多级优化算法求解子问题(44.8-6)和(44.8-7)。一般情况下,总体性约束(如位移约束、总体屈曲约束、频率约束等)和几何形状设计变量划归为系统级,而局部性约束(如应力约束,局部屈曲约束,尺寸变量上、下界等)和尺寸变量应划归为子系统级。

原始优化问题一旦分解成合适的子问题后,将对不同的优化问题应采取不同的优化算法或策略,才会使优化计算更加有效。对多级优化问题式(44.8-6)和式(44.8-7)的求解步骤一般为:

步骤 1: 选择  $Y$  的初值。

步骤 2: 对给定的  $Y$ , 求解第 2 级的第  $i$  个子问题式(44.8-7)。

步骤3: 求解问题 (44.8-6)。

步骤4: 重复步骤2和步骤3直至满足收敛条件。

需要指出的是, 多级优化算法在进行多级分解时一个重要的问题是, 如何考虑不同级优化问题甚至同级不同优化问题间的设计变量与约束的相互耦合影响。某些情况下可以简单地忽略其耦合影响, 而另一些情况下则可能需要通过附加某些约束来考虑。另外, 在具体实施多级优化时, 还可把某级的第  $i$  个子结构处理成第  $i$  个子问题。对各个子问题也可以利用并行算法来求解。

## 4 多学科优化设计

多学科优化设计是有代表性的复杂系统优化问题。目前有许多学者在这领域开展了研究工作。

对于多学科优化设计问题, Batill 等人在并行子空间优化算法基础上提出了多学科设计优化设计并行子空间设计法。并行子空间设计算法 (Concurrent Subspace Design) 简称为 CSD。

多学科优化设计并行子空间设计法包括如下几个主要内容:

### (1) 建立初始数据库, 构造初始响应面

CSD 算法要求选择一组基准设计方案 (Baseline Design), 这些基准设计方案经过系统分析, 可获得相应的系统状态变量。这些设计变量和状态变量构成初始设计数据库, 将这 - 数据库作为样本, 应用三层前馈神经网络来构造系统分析的响应面。

### (2) 子空间设计和优化

CSD 算法允许设计人员依据现有分析工具和设计方法, 各自提出各种设计方案, 分析工具包括计算机软件、统计数据或实验数据。设计方法不仅仅限于优化方法, 还可以是专家系统, 甚至只是设计人员的直觉判断和创造性思维。

为了使各学科的设计人员能有效地影响整个设计活动的进程, 子空间的设计人员可通过改变系统设计变量的一个子集来提出他们的设计方案。任一设计变量可以被一个或几个子空间同时控制, 各子空间能独立并行设计的前提是首先必须解决子空间之间耦合的问题。在 CSD 算法中采用了系统近似分析模型 (响应面) 来解决子空间之间的耦合问题。子空间设计中所需的局部系统状态变量可由该学科已有的分析工具获得, 而子空间设计中所需的非局部状态变量的信息可通过神经网络 (响应面) 来近似地获得。例如, 结构组的设计人员可用结构有限元分析软件获得飞机结构的应力和位移 (局部状态变量), 而在结构分析中所需的气动载荷分布 (非局部状态变量) 不是直接用气动计算模型或吹风实验获得, 而是通过由神经网络构造的

影响面来近似地获得气动载荷分布。CSD 算法中子空间并行设计的结果是由各学科组提出的一组设计方案。

### (3) 更新设计数据库和响应面

将各子空间提出的设计方案进行系统分析, 获得相应的系统状态变量。这些新的设计方案 (设计变量) 及其系统状态变量被加进设计数据库中, 得到一个新的数据库。以这个数据库作为训练三层前馈神经网络的样本, 神经网络经学习后, 获得了一个新响应面。在 CSD 算法中, 总是利用设计数据库的所有数据来训练神经网络, 构造响应面, 因此, CSD 算法中每一迭代过程的设计信息都是被充分利用的。随着设计进程 (迭代过程) 的展开, 数据库越来越丰富, 相应的响应面也越来越精确。

### (4) 系统级优化

CSD 算法中的第四阶段是系统级优化 (或称系统级协调), 其分析模型是采用系统分析的响应面。由于基于响应面的近似系统分析所需的计算量很小, 系统级优化阶段中的优化算法有很大的选择余地。在目前 CSD 算法中, 采用了一种混合优化策略, 即模拟退火算法和梯度下降法的混合搜索策略。

这一阶段的结果是获得一个系统设计方案, 或称为各子空间设计方案的折衷方案。这一方案经系统分析被加进到设计数据库中, 重新训练神经网络后, 得到一个新的响应面, 然后开始新一轮的子空间设计过程。如果相邻两次系统设计方案几乎一致, 则可认为设计过程已收敛, 设计过程结束。

从上述 CSD 算法的过程中看出, CSD 算法的特点可描述为“由于子空间设计和系统级设计所驱动的算法”。因为随着进程的展开 (迭代次数的增加), 每个子空间提出的设计和系统级协调设计的方案逐渐增加, 设计数据库不断丰富, 基于数据库构造的系统近似模型也不断精确。因此, 在于空间设计和系统设计的驱动下, 系统分析的响应面不断进化。另一方面, 子空间和系统级设计从响应面中获得的信息越来越精确, 这个响应面又反过来影响子空间和系统级设计方案。两者相互影响, 相互依赖。最终, 系统级设计方案收敛于优化结果。

下面介绍多学科优化设计 CSD 算法应用于飞机初步设计的实例。飞机设计优化问题可描述为:

寻找一组设计变量:  $K$

最小化目标函数: 飞机起飞重量  $W_0$ 。

并满足性能约束: 最大航程  $Range \geq 8.5$  (mile)

最大平飞速度  $v_{max} \geq 60.0$  (ft/s)

最小平飞速度  $v_{min} \leq 28.0$  (ft/s)

起飞滑跑距离  $S_{L_0} \leq 45.0$  (ft)



最大爬升率  $ROC \geq 7.5$  (ft/s)  
 纵向静稳定裕度大于 15%，小于 25%  
 前轮离地速度  $v_{rotate} \leq 23.0$  (ft/s)

几何约束：机翼和平尾间距大于平均机翼翼根弦长 15%

机翼翼根弦长  $\leq$  机身中段长度

选择其中一些对上述目标函数和约束影响最大的参数作为设计变量，而其余参数则作为固定参数。系统分析的结果包括气动、重量重心、推力特性、性能和操纵等状态变量。CSD 算法中的设计数据库由设计变量和状态变量组成。只选择其中直接与目标函数和约束条件有关的状态变量作为设计数据库的组成部分，这些状态变量称之为最终状态变量，而其余状态变量（如升力系数）称为中间状态变量。这个数据库作为样本数据组，用来训练人工神经网络。神经网络学习后，可获得设计变量和最终状态变量之间的近似映射关系，也就是系统分析的近似模型响应面。

将设计问题分解为四个学科的设计优化问题：

子系统 1 设计优化 (SS1) —— 气动和性能设计优化组

子系统 2 设计优化 (SS2) —— 结构和重量设计优化组

子系统 3 设计优化 (SS3) —— 推进系统设计优化组

子系统 4 设计优化 (SS4) —— 操纵设计优化组  
 每个子系统设计优化的表达形式类似于上述设计优化问题，但每个子系统只负责设计整个设计变量的一个子集，设计变量在各子系统之间可共享。子系统的分析工具包括本学科的分析模型（如气动分析模型）和响应面模型，子系统的设计工具包括优化算法、专家系统或设计人员的经验。在本应用研究中，只有推进子系统应用了专家系统技术。

CSD 算法在初步设计的应用研究包括三部分：① 含 12 个连续型设计变量的设计优化问题。其目的是将 CSD 算法的设计结果能与传统的单级优化算法的结果进行比较，评判 CSD 算法的性能。② 含 29 个连续/离散混合设计变量的设计优化问题。这也是目前 CSD 算法处理的最复杂的实际问题。其目的是检验 CSD 算法能否有效地解决比较复杂的、且含有混合设计变量的实际工程问题，同时为 EPUAV 的多学科并行设计提供一个计算机环境。③ 探讨 CSD 算法是否能有效地处理设计要求（约束）动态性的问题。

表 44.8-1 列出 12 个设计变量的名称、上限和下限和子系统对各设计变量的控制权。设计变量与最终状态变量的关联图列在表 44.8-2 中，此表供构造响应面使用。

表 44.8-1 设计变量情况

序号	设计变量名称	变量	下限	中间	上限	类型	SS1	SS2	SS3	SS4
1	机翼展弦比	ARW	4.5	8.0	10.0	连续	×	×		×
2	机翼面积 (ft <sup>2</sup> )	Sw	6.0	9.0	11.0	连续	×	×		
3	机翼梯形比	taper W	0.5	0.7	1.0	连续	×			
4	机翼后掠角	Sweep W	0.0	5.0	10.0	连续				×
5	机翼扭转角	theta	0.0	2.0	5.0	连续	×			
6	机翼安装角	InC W	-1.0	2.0	8.0	连续	×			
7	平尾展弦比	Arh	1.5	3.5	5.0	连续				×
8	平尾面积/ft <sup>2</sup>	Sh	1.5	2.5	3.5	连续		×		×
9	机身中段长度	Lm	0.75	1.5	1.75	连续	×	×		×
10	机翼的位置	Xwing	0.3	0.75	1.2	连续				×
11	螺旋桨直径/in	D	11.0	12.0	13.0	连续		×	×	
12	螺距/in	Pitch	5.0	6.0	7.0	连续			×	

注：SS1——气动和性能设计优化组；SS2——结构重量设计优化组；SS3——推进系统设计优化组；SS4——操纵设计优化组；未注明的长度单位为英尺。

#### (1) 单级优化的结果

基于系统分析模型，采用约束变尺度法对 12 个设计变量进行优化。CVM 是一种比较优秀的基于梯度的优化算法，其可靠性已经过严格的测试。由于初始点的选取对最终优化结果会有影响，我们选择了三种情况进行设计优化。第一种情况的初始点选为所有设计变量为中间值，第二、三种情况的初始点分别靠近设计

变量的上限、下限。表 44.8-4 中的情况 1、2、3 分别总结了优化设计的结果。从表 44.8-2 中看出，三次优化均在可行域内获得了设计点。不同的初始点得到不同的局部最优解，且所需的系统分析次数均比较多。从不同的初始点出发，得到优化结果所需的系统分析次数相差很大。

#### (2) CSD 算法的结果

根据 CSD 算法, 首先需建立初始设计数据库。为了讨论初始设计数据库对优化结果的影响, 我们选取了二个初始设计数据库。首先将各设计变量的取值范围分成四等份, 每个设计变量由小到大依次取值, 而其余设计变量取中值, 组成初始设计点, 共有 49 个设计点。设计点经系统分析后, 得到最终状态变量。这些设计

点和相应的状态变量构成一个初始设计库。另一个设计数据库的构成方法是: 每个设计变量依次取上限、中值、下限, 而其余设计变量取中值, 组成初始设计点, 共有 25 个初始设计点, 经系统分析后, 便获得第二个初始设计库。两个初始设计数据库的统计特征列在表 44.8-3 中, 只有一个设计点在可行域内。

表 44.8-2 设计变量与最终状态变量的关联表 (12 个设计变量)

序号	设计变量	$W_{to}$	$U_{n,ax}$	$U_{n,min}$	ROC	$S_{to}$	$R_{max}$	静稳定裕度	$V_{爬爬}$
1	ARW	×	×	×	×	×	×	×	×
2	Sw	×	×	×	×	×	×	×	×
3	taper W		×	×	×	×	×	×	×
4	Sweep W		×	×	×	×	×	×	×
5	theta		×	×	×	×	×	×	×
6	InC W		×	×	×	×	×		×
7	Arh		×	×	×	×	×	×	×
8	Sh	×	×	×	×	×	×	×	×
9	Lm	×	×	×	×	×	×	×	×
10	Xwing		×	×	×	×	×	×	×
11	D	×	×	×	×	×	×	×	×
12	Pitch		×	×	×	×	×		

表 44.8-3 12 个设计变量的初始设计数据库特征

初始设计点数	平均 $W_{to}/lb$	可行设计点数目	可行域内最小 $W_{to}/lb$	初始设计点中最小 $W_{to}/lb$
25	6.432	1	6.45	6.125
49	6.258	1	6.25	5.960

应用 CSD 算法, 对设计问题进行了三次设计优化。在第一、二次优化中, 采用了第一个初始设计数据库, 但神经网络的初始权值不同, 第三次优化中, 采用了第二个初始设计数据库 (25 个设计点)。最后的优化结果分别列在表 44.8-4 中的情况 4、5 和 6。从表中可看出, 三次优化均在可行域内找到最优解, 三次优化结果很接近, 而且所需的系统分析次数也相近。

根据表 44.8-4, 比较单级优化算法与 CSD 算法的优化结果, 可看出: ①CSD 算法的结果明显好于单级优化算法的结果; ②CSD 算法所需的系统分析次数至

少比单级优化算法少于 50% (比较情况 2、3 和情况 4、5、6), 甚至于至少一个数量级 (比较情况 1 和情况 4、5、6); ③CSD 算法的稳健性也好于单级优化算法, 以不同的初始值 (包括不同的初始设计数据库和神经网络的初始权值) 出发, CSD 算法获得相近的优化结果; 而单级优化算法, 从不同初始点出发, 获得了不同的优化结果。

表 44.8-4 12 个设计变量 EPUAV 问题的优化结果

序号	方法	初始设计数据库	优化结果 $W_{to}/lb$	总系统分析次数
1	NAND	—	5.905	887
2	NAND	—	5.732	190
3	NAND	—	5.735	271
4	CSD	49	5.709	101
5	CSD	49	5.708	83
6	CSD	25	5.706	82

# 第 45 篇 可靠性设计

主 编 谢里阳 王永岩  
编写人 谢里阳  
王永岩  
李 佳

# 第1章 产品可靠性及可靠性设计

## 1 产品质量与可靠性

### 1.1 产品质量

产品质量是产品满足使用要求所具备的固有属性,其中既包括功能指标,也包括可靠性指标。这里的“产品”可以指作为设计对象的任何零件、部件、设备或系统。产品的功能是由产品的技术指标决定的。功能指标是产品的基本指标,如果功能指标不满足要求,产品就没有应用价值。另一方面,即使产品的各项技术指标都很先进,如果不可靠,同样也没有使用价值。产品的功能能否得到发挥,在很大程度上取决于产品的可靠性水平。

可靠性是产品(零件或系统)的质量指标,表示产品正常工作的概率,有明确的定义。然而,可靠性有时也被用于在一般的意义上概括地表示可靠性(存活率)、可用性、可维修性和耐久性。可用性与存活率不同,是系统可用于工作的时间的比例,可用性通常应用于过程装备等连续工作的系统;而可靠性则多用于车辆等非连续作业的场所。可维修性涉及恢复系统到工作状态所花费的时间,包括预防性维修或计划维修时间及故障修复时间。耐久性用于诸如磨损和疲劳等时间相关的失效模式。这四个指标是相互关联的,其中任何一个指标的改变都将对其他指标产生影响。对这些指标的描述将需要用到数学知识,包括不确定性、零件与系统之间的关系等等。

在可靠性问题中,有一点值得注意的是,机械系统及零件的可靠性问题与电器元件的可靠性问题有明显的不同。例如,电子元件的寿命一般服从指数分布,失效率为常数;机械零件的寿命则多用 Weibull 分布(威布尔分布)或对数正态分布描述,失效率随时间变化。电子系统中各元件的失效通常可以假设是相互独立的,可以用简单的串、并联系统模型建立系统失效概率与元件失效概率之间的关系;而机械系统中各零件的失效存在显著的相关性,若沿用简单的串、并联系统模型建立系统失效概率与元件失效概率之间的关系,则可能导致严重的误差。

### 1.2 产品可靠性与有效性

产品的可靠性指产品在规定的条件下,规定的时间内完成规定功能的能力。规定的条件不同,产品的可

靠性也将不同。例如,同一台设备在室内、野外(寒带或热带,干燥地区或潮湿地区)、海上、空中等不同的环境条件下工作,其可靠性也是不同的。

“规定的时间”是可靠性区别于产品其他质量属性的重要特征,产品的可靠性水平会随着使用或贮存时间的增加而降低。因此以数学形式表示的可靠性特征量是时间的函数。这里的时间概念不限于一般的时间概念,也可以是产品操作次数、载荷作用次数、运行距离等。

“规定功能”是要明确具体产品的功能是什么,以及怎样才算是完成规定功能。产品丧失规定功能称为失效,对可修复产品通常也称为故障。

机械产品一般是可维修的,要使一台设备发挥更好的作用,不仅要求在单位时间内出现的故障次数少,故障间隔时间长,而且要求维修时间短。将产品的能工作时间与总时间之比称为产品的有效性,产品的有效性是指可修产品维持其功能的能力。

### 1.3 产品全寿命周期费用

质量包括产品的性能指标、专门特性、经济性、适应性等多个方面。产品的性能指标是描述其基本功能的参数,如结构的强度、发动机的输出功率等。专门特性描述保持规定性能指标的能力,包括产品的可靠性、维修性、保障性、安全性、测试性等,如结构强度在规定时间内不发生退化、发动机能连续工作若干小时并保证在此期间输出功率不低于规定的值。经济性(全寿命周期费用)指在整个寿命期内的总费用。

随着现代系统的复杂化,专门特性显得更加重要。这是由于:

1) 工程系统日益庞大和复杂,带来了系统可靠性和安全性下降、风险概率增加的问题;如航天飞机作为一个由数十万个零件组成的系统,其可靠性是至关重要的问题。

2) 应用环境更加复杂和恶劣。从深海到太空,严酷的环境对系统高可靠性、高安全性等综合特性的实现提出了新的挑战。

3) 系统要求的持续无故障任务时间加长。如太空探测器的长时间无故障飞行要求、通信网络的关键任务不停机要求等,迫使工程系统必须具有良好的可靠性、维修性等专门特性。

4) 系统的专门特性与使用者的生命安全直接相

关。如核能系统、载人航空航天器、高速列车等系统的可靠与安全是生命安全的基本保证,受到了强烈的关注。

5) 市场竞争的影响。“性能优良、功能齐全”并不是用户选择产品时考虑的惟一因素,产品是否可靠、是否好修、使用维护保养费用多少、寿命多长都对用户的选择产生重要影响。

系统优化权衡的核心是效能、寿命周期费用两个概念之间的权衡。系统效能是系统在规定的条件下满足给定特征和服务要求的能力。系统的寿命周期费用(Life Cycle Costs,简称LCC),是指在系统的整个寿命周期内,为获取并维持系统的运行(包括处置)所花费的总费用。它包括硬件、软件的研制费、生产费、后勤保障费用,以及在研制、采办、使用、技术保障和处置过程中所需的各种费用。不同的系统,其寿命周期费用构成不完全相同,各构成成分间的比例关系也不完全一样。

在讨论产品的可靠性时,还应该注意产品的可靠性与成本和利润三者之间的关系。强调产品的质量,并不是说产品的质量越高越好,而是在满足使用要求的前提下,尽可能保持质量、效率与费用这三个基本目标间的平衡。图 45.1-1 为产品的可靠性与产品的设计、制造成本及使用维护费用之间的一般关系。由图可见,提高产品的设计可靠性,会导致生产成本的增加,但使用、维护成本随着可靠性的提高而降低。图中总费用曲线是各项成本及费用之和。

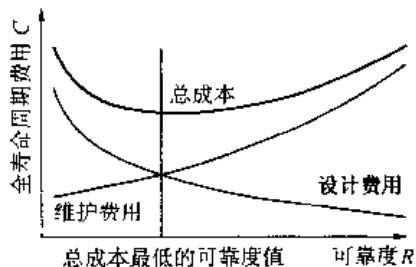


图 45.1-1 产品全寿命周期费用与可靠性之间的关系曲线

## 1.4 可靠性设计

产品可靠性设计是指在产品的开发设计阶段将载荷、强度等有关设计量及其影响因素作为随机变量对待,应用可靠性数学理论与方法,使所设计的产品满足预期的可靠性要求。产品开发设计阶段的主要内容还包括预测设计对象的可靠度、找出并消除薄弱环节、不同设计方案之间的可靠性指标比较等,可分为定量分析与定性分析两个方面。

定量分析是应用概率统计方法、布尔代数、马尔可夫过程理论、故障树分析等计算产品的失效概率。

定性分析是通过故障模式、影响及致命度分析

(FMECA)、事件树分析(ETA)、故障树分析(FTA)等对事故种类、原因、后果等进行罗列和分析。

## 2 可靠性问题的统计描述和表达

### 2.1 直方图和密度函数

产品(零件或系统)的可靠性特征量可以通过各种统计方法获得,可以用函数或图形很直观地表示。用图形描述可靠性特征的最简单的方法是失效频度直方图。

如图 45.1-2a 中所列某产品的失效时间(寿命)是在一定的时间范围内随机出现的。图中哪个时间段中的试验数据点靠得愈密,说明失效在这个时间段发生的越频繁。为了直观地表达,可以作出失效频度直方图。其步骤为:

1) 先把  $n$  个离散的失效时间(产品寿命)观测值  $t_i$  ( $i=1\sim n$ )按由小至大顺序排列,即  $t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_n$ ,再将横坐标划分成时间区间,并求出落在各区间中的失效数(样本数)。如果某次失效正好落在边界上,就各算一半给两个相邻的区间。适当选择区间大小或边界一般可以避免这种情况发生。

将时间轴划分成区间并把失效时间归类到相应区间称为对观测值分级。在分级之后,若在每一级中的失效时间观测值都用同一数值—该区间的中值来代替。由原来不同的数值变成了区间的中值,显然遗失了一些细节信息。但通过分级,以信息的损失为代价获取了表达的直观性。

可见,区间的数量如果选得过宽,就会遗失过多的信息。相反,如果区间选得过细,则会在某些区间出现空档,使连续的失效分布中断,因而不适宜于确切的描述。区间数  $k$  的粗略估计可以通过下式计算

$$k \approx \sqrt{n} \quad (45.1-1)$$

2) 确定区间宽度。先从观测值中找出最大值和最小值,求出总区间范围。一般取各区间为等宽,故区间宽度计算公式为:

$$h = (t_n - t_1) / k \quad (45.1-2)$$

3) 确定区间边界,取不大于观测值中最小值  $t_1$  的一个合适的坐标值为起点,由该点起每增加一个区间宽度处作为一个区间边界,直至全部观测值都被包含为止,并取各区间中值作为该区间的代表值。

4) 作频数和频率分布表,绘频数和频率直方图(见图 45.1-2b)。各级(区间)中的样本数(失效数)用不同高度的长条图形来表示,长条的高度或纵坐标值可以是绝对频度  $h_{abs}(i)$

$$h_{abs}(i) = \text{区间 } i \text{ 中的样本数} \quad (45.1-3)$$

或常用的相对频度  $h_{rel}(i)$

$$h_{rel}(i) = \frac{\text{区间 } i \text{ 中样本数}}{\text{样本总数 } n} \quad (45.1-4)$$

在图 45.1-2b 中,长条图形的高度为相对频度,纵坐标以百分数表示。

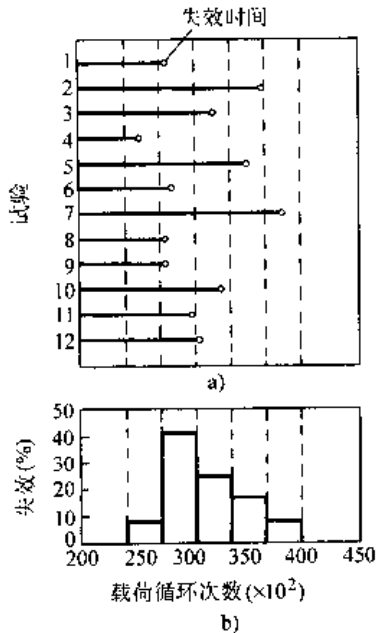


图 45.1-2 失效时间和失效频度直方图  
a)失效时间 b)失效频度直方图

也可以用“经验密度函数  $f^*(t)$ ”取代直方图来描述失效特征(图 45.1-3)。做法是把直方图中长条图形的中点用直线连接起来表示失效频度与失效时间之间的关系。“经验”是指这个密度函数是在有限的失效数据(样本)基础上作出来的。

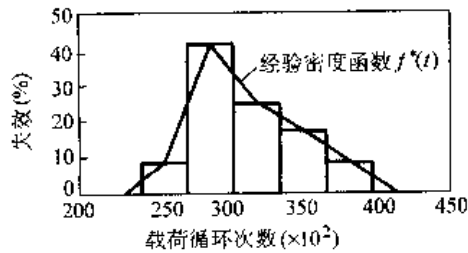


图 45.1-3 失效频度直方图和经验  
概率密度函数  $f^*(t)$

如果试验零件的数量不断增加,就可以得到真正的密度函数。区间的数量按照式(45.1-1)随试验数据的增加而增加,这意味着,当得出的频度纵坐标值基本保持不变时,区间将取得愈来愈缩小。对极限过程  $n \rightarrow \infty$ ,直方图的轮廓逼近于一条光滑连续曲线(图 45.1-4)。这条极限曲线表示的是真正的密度函数  $f(t)$ 。

极限  $n \rightarrow \infty$  意味着,对一个很大总体的所有零件都进行试验,以此得到精确的失效特征。基于实有的试验规模总是只能得到经验密度函数  $f^*(t)$ 。在试验数

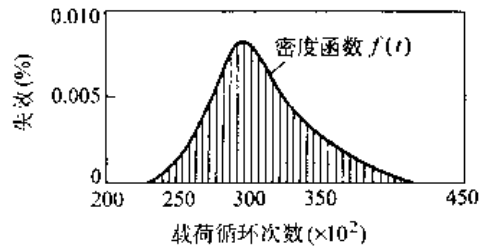


图 45.1-4 失效频度直方图与概  
率密度函数  $f(t)$

据很少的情况下,经验密度同真实密度函数会有显著  
的差异。

频度直方图或密度函数把失效数描述为时间的函数。它们是描述失效特征最简单而又直观的途径。通过频度直方图,除了可以表示失效时间的离散范围以外,还能表达大多数失效出现的范围。

## 2.2 分布函数和失效概率

在许多情况下,并不是对某个时刻或某个区间内的失效数感兴趣,而更多的是需要知道,在某个时刻(或区间)之前总共共有多少零件失效。这个指标可以用累积频度直方图来表示。将如图 45.1-5a 所示的直方图中各区间的失效数对失效时间进行累加,就可构成图 45.1-5b 所示的累积频度直方图。

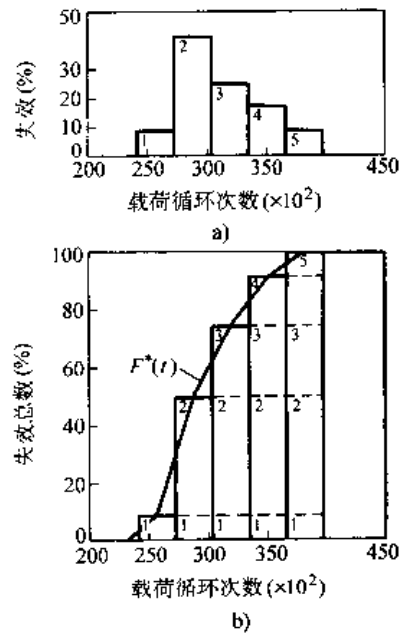


图 45.1-5 累积频度和经验分布函数  
a) 频度直方图

b) 累积频度直方图和经验分布函数  $F^*(t)$

由此,可以得到第  $m(m=1 \sim k)$  级累积频度  $H(m)$

$$H(m) = \sum_{i=1}^m h_{mi}(t) \quad (45.1-5)$$

同密度函数一样,到某时刻的失效数的总和也可以用 一个函数来表示。这个函数被称为“经验分布函数  $F^*(t)$ ”(图 45.1-5b)。

如果将试件数量不断增加,区间选得愈来愈小,在极限  $n \rightarrow \infty$  情况下,累积频度直方图的轮廓逼近于一条光滑曲线,即可得到真正的分布函数  $F(t)$ (图 45.1-6)。

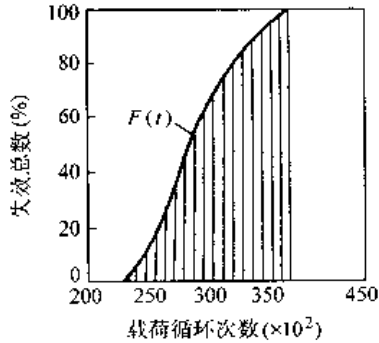


图 45.1-6 累积频度和分布函数

分布函数总是从  $F(t)=0$  开始,然后单调增加。因为对于任意时刻或区间,所观察到的失效频度总是正值。当所有零件都失效时,分布函数终止于  $F(t)=1$ 。

由式(45.1-5)并考虑到极限过程可得出分布函数作为密度函数的积分

$$F(t) = \int_0^t f(t) dt \quad (45.1-6)$$

由此可知密度函数为分布函数的导数

$$f(t) = \frac{dF(t)}{dt} \quad (45.1-7)$$

在可靠性理论中,一般将分布函数  $F(t)$  称为“失效概率”,因为函数  $F(t)$  描述了到某一时刻发生失效的概率。

### 2.3 可靠度

可靠度是产品在规定条件下和规定时间内,完成规定功能的概率,记为  $R(t)$ 。可靠度是时间的函数,故  $R(t)$  也称为可靠度函数。若产品寿命  $t$  的概率密度为  $f(t)$ ,可靠度函数可用公式表示如下

$$R(t) = \int_t^{\infty} f(t) dt \quad (t \geq 0, 0 \leq R(t) \leq 1) \quad (45.1-8)$$

与之对应,产品失效概率  $F(t)$  定义为

$$F(t) = \int_0^t f(t) dt \quad (45.1-9)$$

显然,  $R(t) + F(t) = 1$ 。

可靠度、失效概率的统计意义可表述如下。设有  $n$  个同一型号的产品(概率意义上相当于属于同一母体),工作到时刻  $t$  时有  $n(t)$  个失效,则

$$R(t) \approx \frac{n - n(t)}{n} \quad (45.1-10)$$

$$F(t) \approx \frac{n(t)}{n} \quad (45.1-11)$$

将失效函数  $F(t)$  对时间  $t$  微分,即为失效密度函数  $f(t)$ (也叫故障密度函数)

$$f(t) = \frac{dF(t)}{dt} = - \frac{dR(t)}{dt} \quad (45.1-12)$$

$f(t)$  的统计意义可表达为

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{n(t + \Delta t) - n(t)}{n \Delta t} \quad (45.1-13)$$

到某一时刻仍具有工作能力产品的多少可以用可靠频度直方图来表示(图 45.1-7)。将已经失效的产品从产品总数中减掉,就得到这个直方图。在图 45.1-7 中将长条的中点用线段连接起来,表示经验可靠度。

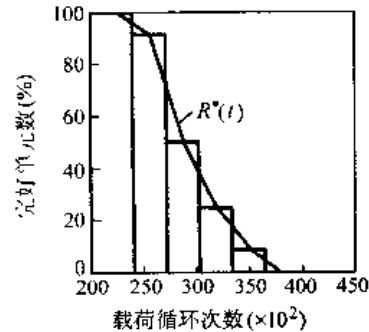


图 45.1-7 可靠频度直方图及经验可靠度

可靠度  $R(t)$  是失效概率  $F(t)$  的补函数

$$R(t) = 1 - F(t) \quad (45.1-14)$$

由于在  $t=0$  时还没有失效出现,可靠度  $R(t)$  始终从  $R(t)=100\%$  开始。然后,函数  $R(t)$  单调下降,当所有单元都失效时终止于  $R(t)=0$ 。

在可靠性理论中,可靠度  $R(t)$  也称之为“可靠性  $R(t)$ ”。用函数  $R(t)$  可以客观和定量地确定通常只是定性使用的“可靠性”概念。

### 2.4 失效率

失效率也称故障率,定义为工作到时刻  $t$  时尚未失效的产品,在时刻  $t$  以后的单位时间内发生失效的概率。失效率一般记为  $\lambda$ ,它也是时间  $t$  的函数,因此也记为  $\lambda(t)$ ,称为失效率函数,有时也称为故障率函数或风险函数。

根据定义,失效率是在时刻  $t$  尚未失效的产品在  $t + \Delta t$  的单位时期内发生失效的条件概率,即

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t < T \leq t + \Delta t | T > t) \quad (45.1-15)$$

其观测值为在时刻  $t$  以后的单位时间内发生失效的产品数与工作到该时刻尚未失效的产品数之比

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{n(t + \Delta t) - n(t)}{[n(t) \Delta t]} \quad (45.1-16)$$

例如,100个产品工作到80h时尚有50个仍未失效,在80~81h内又失效2个,则  $\Delta n_i(t) = 2, n_i(t) = 50, \Delta t = 1$ ,故

$$\hat{\lambda}(80) = \frac{2}{50 \times 1} = 0.04$$

平均失效率是指在某一规定时期内失效率的平均值。如图45.1-8所示,在  $(t_1, t_2)$  内失效率的平均值

$$\bar{\lambda}(t) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \lambda(t) dt \quad (45.1-17)$$

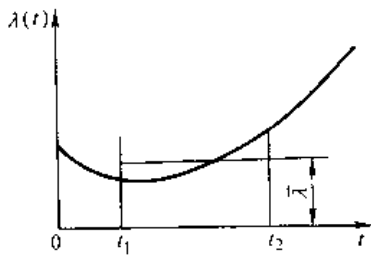


图 45.1-8 失效率及其在有效寿命期间的均值

平均失效率的观测值,对于不可修复的产品是指在一个规定的时期内失效数与累积工作时间之比。对于可修复的产品是指它的使用寿命内的某个观测期间一个或多个产品的故障发生次数与累积工作时间之比。即

$$\hat{\lambda}(t) = \frac{r}{\Sigma t} \quad (45.1-18)$$

式中  $r$ ——在规定时间内失效数;

$\Sigma t$ ——规定时间内的累积工作时间。

失效率的单位用单位时间的百分数表示,例如%/10<sup>3</sup>h,可记为10<sup>-5</sup>/h。对高可靠性则用10<sup>-6</sup>/h为单位。失效率的单位为单位时间的失效数,即1/h、1/km、1/次等。

例如失效率  $\lambda = 0.0025 / (10^3 h) = 0.25 \times 10^{-5} / h$ ,表示10万个产品中,每小时只有0.25个产品失效。

失效率  $\lambda(t)$  与可靠度  $R(t)$ 、失效概率密度函数  $f(t)$  等有以下关系

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (45.1-19)$$

$$R(t) = e^{-\int_0^t \lambda(t) dt} \quad (45.1-20)$$

若  $\lambda(t)$  为常数,则有

$$R(t) = e^{-\lambda t} \quad (45.1-21)$$

工程中,失效率  $\lambda(t)$  可以按下式计算

$$\lambda(t) = \frac{\text{(在时刻 } t \text{ 或级 } i \text{ 后的单位时间内的) 失效数}}{\text{(在时刻 } t \text{ 或级 } i \text{) 完好产品总数}} \quad (45.1-22)$$

图45.1-9所示为失效率直方图和经验失效率  $\lambda^*(t)$  曲线。值得注意的是,最末级的失效率必然趋向

于 $\infty$ ,这是由于不可存在完好产品。

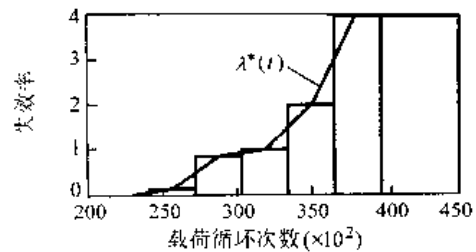


图 45.1-9 失效率直方图和经验失效率  $\lambda^*(t)$  曲线

失效率  $\lambda(t)$  可以从密度函数  $f(t)$  和可靠度  $R(t)$  这两个函数的商求得

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (45.1-23)$$

图45.1-10所示失效时间为  $t_x$  时式(45.1-23)的图示。

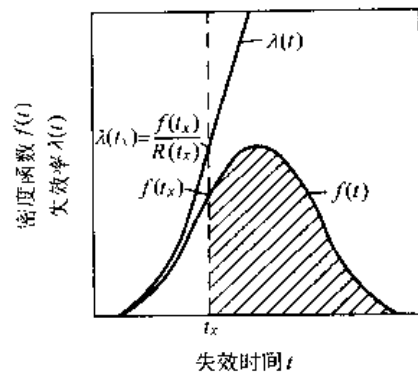


图 45.1-10 概率密度函数、可靠度与失效率

某时刻  $t$  的失效率可以解释为,在时刻  $t$  现有零件完好的前提下,一个零件失效的危险程度。倘若考察一个确定的时刻  $t$ ,失效率就表明,在下一个单位时间内,当前所有完好零件中将以多大的比率失效。

失效率  $\lambda(t)$  很常用,不仅用于如图45.1-9所描述的疲劳失效,而且还用于所谓的早期失效和偶然失效。典型的失效率曲线如图45.1-11,传统上根据曲线的形状把它称为“浴盆曲线”。浴盆曲线可以明显地划分为三个范围:早期失效范围、偶然失效范围以及耗损失效范围。

早期失效范围内的失效率表现为随时间下降的趋势。零件的失效危险随着时间的推移而下降。引起早期失效的原因主要是由于装配错误、制造缺陷、材料缺陷或明显的设计错误。

在偶然失效范围内失效率保持为常数,因而零件失效的危险性保持不变。通常,这个危险性相对地低。这种偶然失效是由操作或维护缺陷等随机因素引起的。一般地,这种失效难以事先预料。

在耗损失效范围内失效率随时间急剧增大。零件失效的危险性随着时间的增加愈来愈大,乃至相当高。



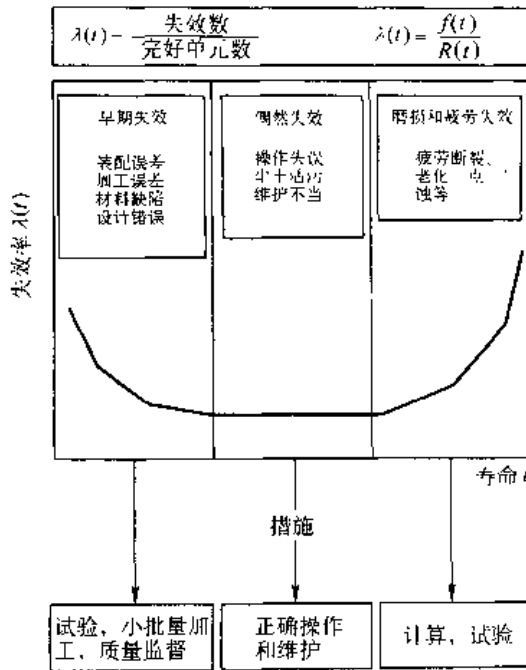


图 45.1-11 失效率曲线(浴盆曲线)

图 45.1-12 表示机械零件的失效曲线,由图可见,机械零件的失效率曲线没有  $\lambda = \text{常数}$  的部分。因为机械零件的主要失效形式如疲劳、腐蚀、磨损及蠕变等都属于典型的损伤累积失效,而且影响失效的随机因素很复杂,所以,随着时间的延长,失效率是递增的,即  $\lambda(t)$  为非降函数。在试验和运行初期,少数零件由于材料本身存在的严重缺陷,或者工艺过程中形成的裂纹和缺陷,以及严重的应力集中等,使得少数零件一旦承受载荷就很快失效,因而出现一定的初期失效率,但和电子元件相比,它要小得多。随后,零件进入正常使用期,由于损伤不断积累,失效率不断增大。

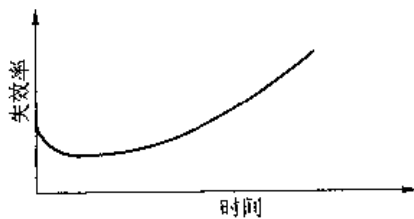


图 45.1-12 机械零件的失效率曲线

### 3 统计学参数

产品寿命等可靠性特征量可以用前面介绍的函数描述。在许多情况下,知道有关参数(例如失效时间或寿命)的“中值”位于何处以及个体与均值的“离散”程度即可满足需要。对此,可以使用所谓的“位置参数和离散参数”(最常用的参数是均值和标准差)。当然,用参数来表征随机变量是一种简化,会遗失掉一些信息。

#### (1) 均值

通常简称为均值的经验算术平均值由失效时间  $t_1, t_2, \dots, t_n$  按下式求出

$$t_m = \frac{t_1 + t_2 + \dots + t_n}{n} = \frac{1}{n} \sum_{i=1}^n t_i \quad (45.1-24)$$

算术均值对与其偏离较大的“离散值”很敏感,也就是说,一个极短的或极长的失效时间会大大地影响均值的大小。

#### (2) 方差

经验方差  $s^2$  描述了个体与算术均值的平均偏差,它衡量个体与均值  $t_m$  的离散程度

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (t_i - t_m)^2 \quad (45.1-25)$$

#### (3) 标准差

经验标准差  $s$  通过对方差开方得到

$$s = \sqrt{s^2} \quad (45.1-26)$$

标准差与方差相比的优越之处在于,它具有与失效时间相同的量纲。

#### (4) 中位数

中位数是当失效数正好为总失效数一半时的失效时间。因此,中位数  $t_{\text{median}}$  可以简单地通过失效概率  $F(t)$  与时间的关系求出。

$$F(t_{\text{median}}) = 0.5 \quad (45.1-27)$$

如果失效状态用密度函数  $f(t)$  表示,那么中位数将曲线  $f(t)$  与横坐标轴围成的面积分成相等的两部分。

中位数与均值相比最大的优点在于,它对偏离较大的“离散值”很不敏感。一个很小的或很大的失效时间都不会使中位数移动。

#### (5) 众数

最常出现的数值称为众数。因此,众数  $t_{\text{modal}}$  可以从密度函数  $f(t)$  中简单求出:  $t_{\text{modal}}$  是当密度函数为最大时的失效时间。

在概率论中,众数具有重要的意义。假如进行一个试验,就可以预料,大多数零件在寿命等于众数时失效。

均值、中位数和众数在一般的非对称分布时各不相同(图 45.1-13)。只有当密度函数曲线完全对称时

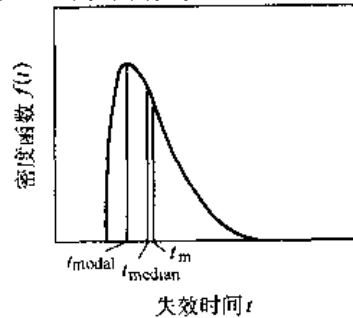


图 45.1-13 概率密度分布及均值、中位数和众数

(正态分布就属于这种情形)这些参数值才相等。

### 4 可靠性问题示例

下面用一个拉压交变载荷作用下的轴进行疲劳循环试验的实例详细示范前面介绍的函数和参数的估计方法与步骤。

#### (1) 试验结果

在应力幅为 380MPa 时,用 20 根轴进行试验,得到如下失效时间(单位:h):

100000, 90000, 59000, 80000, 126000, 117000, 177000, 98000, 158000, 107000, 125000, 118000, 99000, 186000, 66000, 132000, 97000, 87000, 69000, 109000

作为评价的第一步,将试样的失效时间按从小到大顺序排列:

$t_1 = 59000, t_2 = 66000, t_3 = 69000, t_4 = 80000,$   
 $t_5 = 87000, t_6 = 90000, t_7 = 97000, t_8 = 98000,$   
 $t_9 = 99000, t_{10} = 100000, t_{11} = 107000, t_{12} = 109000,$   
 $t_{13} = 117000, t_{14} = 118000, t_{15} = 125000, t_{16} =$   
 $126000, t_{17} = 132000, t_{18} = 158000, t_{19} = 177000, t_{20}$   
 $= 186000$

#### (2) 分级

级数由近似公式(45.1-1)得出

$$k = \sqrt{n} = \sqrt{20} = 4.5$$

由此,应选择 4 或 5 级。为得到较好的结果而分成 5 级。

级宽(区间宽度) $h$  由最长与最短失效时间的差除以级数  $k$  得到

$$h = \frac{t_{20} - t_1}{k} = \frac{186000 - 59000}{5} = 26000$$

从最短失效时间开始,得到以下各级的寿命范围:

- 级 1: 59000, ..., 85000
- 级 2: 85000, ..., 111000
- 级 3: 111000, ..., 137000
- 级 4: 137000, ..., 163000
- 级 5: 163000, ..., 189000

#### (3) 密度函数

根据式(45.1-4),可以由各级的失效数量得出相对频度:

- 级 1: 4 根失效;  $h_{rel,1} = 4/20 = 20\%$ ,
- 级 2: 8 根失效;  $h_{rel,2} = 8/20 = 40\%$ ,
- 级 3: 5 根失效;  $h_{rel,3} = 5/20 = 25\%$ ,
- 级 4: 1 根失效;  $h_{rel,4} = 1/20 = 5\%$ ,
- 级 5: 2 根失效;  $h_{rel,5} = 2/20 = 10\%$ 。

用这些值可以作出失效频度直方图和经验密度函数  $f^*(t)$ (图 45.1-14)。

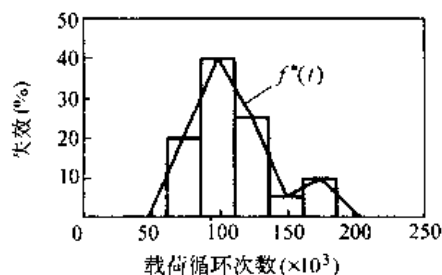


图 45.1-14 轴寿命试验频度直方图和经验密度函数

#### (4) 失效概率

根据式(45.1-5)将失效频度相加得到累积频度直方图和经验失效概率  $F^*(t)$ :

级 1: 累积频度  $H_1 = h_{rel,1} = 20\%$ ,

级 2: 累积频度  $H_2 = H_1 + h_{rel,2} = 20\% + 40\% = 60\%$ ,

级 3: 累积频度  $H_3 = H_2 + h_{rel,3} = 60\% + 25\% = 85\%$ ,

级 4: 累积频度  $H_4 = H_3 + h_{rel,4} = 85\% + 5\% = 90\%$ ,

级 5: 累积频度  $H_5 = H_4 + h_{rel,5} = 90\% + 10\% = 100\%$ 。

累积频度直方图和经验失效概率  $F^*(t)$  如图 45.1-15 所示。

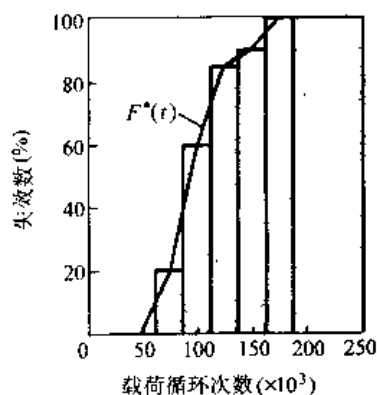


图 45.1-15 轴寿命试验累积频度直方图和经验失效概率  $F^*(t)$

#### (5) 可靠度

根据与失效概率的关系,由式(45.1-14)很容易地可以得出可靠度:

级 1: 可靠度  $R^*_1 = 100\% - H_1 = 100\% - 20\% = 80\%$ ;

级 2: 可靠度  $R^*_2 = 100\% - H_2 = 100\% - 60\% = 40\%$ ;

级 3: 可靠度  $R^*_3 = 100\% - H_3 = 100\% - 85\% = 15\%$ ;

级 4: 可靠度  $R^*_4 = 100\% - H_4 = 100\% - 90\% = 10\%$ 。

10%;

级 5: 可靠度  $R^* = 100\% - H = 100\% - 100\% = 0\%$ ,

可靠度直方图和经验可靠度  $R^*(t)$  如图 45.1-16 所示。

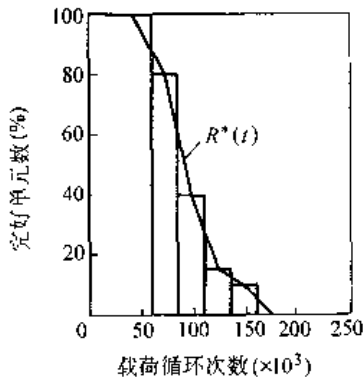


图 45.1-16 可靠度直方图和经验可靠度

(6) 失效率

为求失效率可以用已经求出的相对失效频度和可靠度。由这两个值的商得出失效率:

- 级 1: 失效率  $\lambda_1 = h_{rel,1} / R^*_1 = 20\% / 80\% = 0.25$ ,
- 级 2: 失效率  $\lambda_2 = h_{rel,2} / R^*_2 = 40\% / 40\% = 1.00$ ,
- 级 3: 失效率  $\lambda_3 = h_{rel,3} / R^*_3 = 25\% / 15\% = 1.67$ ,
- 级 4: 失效率  $\lambda_4 = h_{rel,4} / R^*_4 = 5\% / 10\% = 0.50$ ,
- 级 5: 失效率  $\lambda_5 = h_{rel,5} / R^*_5 = 10\% / 0\% = \infty$ 。

这些失效率如图 45.1-17 所示。

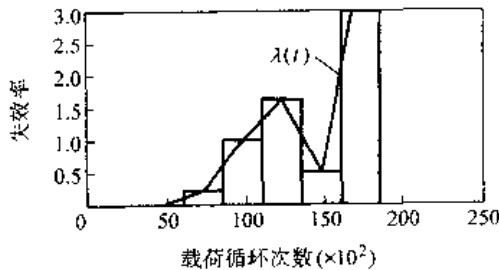


图 45.1-17 失效率

(7) 均值、中位数和众数(位置参数)

经验算术均值为

$$t_m = \frac{t_1 + t_2 + \dots + t_n}{n} = \frac{59 + 66 + \dots + 186}{20} \times 10^4 = 110000$$

中位数可以最简单的由图 45.1-15 中的经验失效概率  $F^*(t)$  求出。它从经验失效概率线和 50% 累积频度线的交点求出。轴寿命的中位数为:  $t_{median} = 95000$ 。

众数相应于密度函数为最大时的失效时间, 因此可以从图 45.1-14 中得出。试验轴的众数为:  $t_{moda} = 98000$ 。

(8) 方差和标准差(离散参数)

试验系列的经验方差用式(45.1-11)算出

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (t_i - t_m)^2 = \frac{1}{19} [(59 - 110)^2 + (66 - 110)^2 + \dots + (186 - 110)^2] \times 10^8 = 1170400$$

对方差求根得到标准差

$$s = \sqrt{s^2} = 34200$$

## 5 产品可靠性指标

### 5.1 平均寿命

在产品的寿命指标中, 最常用的是平均寿命。对于不可修复的产品, 平均寿命是指产品从开始使用到失效这段有效工作时间的平均值, 记为 MTTF (Mean Time To Failure)。对于可修复的产品, 平均寿命指的是平均无故障工作时间, 记为 MTBF (Mean Time Between Failures)。

平均寿命  $\theta$  与其他有关参数的关系

$$\theta = \int_0^{\infty} t f(t) dt \quad (45.1-28)$$

$$\theta = \int_0^{\infty} R(t) dt \quad (45.1-29)$$

当产品失效率  $\lambda(t)$  为常数  $\lambda$  时, 有

$$\theta = 1/\lambda \quad (45.1-30)$$

### 5.2 寿命方差与标准差

平均寿命是一批产品中各个产品的寿命的平均值。它只能反映这批产品寿命分布的中心位置, 而不能反映各产品寿命与此中心位置的偏离程度。

寿命方差和标准差是用来反映产品寿命离散程度的特征值。寿命方差为

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (t_i - \theta)^2 \quad (45.1-31)$$

寿命标准差为

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (t_i - \theta)^2} \quad (45.1-32)$$

### 5.3 可靠寿命、中位寿命与特征寿命

可靠寿命是指可靠度为给定值  $R$  时的工作寿命, 用  $t_R$  表示。

中位寿命是指可靠度  $R=50\%$  时的工作寿命, 用  $t_{0.5}$  表示。

特征寿命是指可靠度  $R=e^{-1}$  时的工作寿命, 用  $t_{e^{-1}}$  表示。

可靠寿命是给定的可靠度所对应的时间, 一般记

为  $t(R)$ 。

如图 45.1-18 所示,一般可靠度随着工作时间  $t$  的增大而下降,对给定的不同  $R$ ,则有不同  $t(R)$ ,即

$$t(R) = R^{-1}(R) \quad (45.1-33)$$

式中  $R^{-1}$  ——  $R$  的反函数,即由  $R(t) = R$  反求  $t$ 。

可靠寿命的观测值是能完成规定功能的产品的比例恰好等于给定可靠度时所对应的时间。

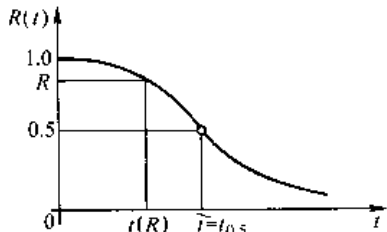


图 45.1-18 可靠寿命和中位寿命

**例** 某产品的失效率为常数,即  $\lambda(t) = \lambda = 0.25 \times 10^{-4} \text{h}^{-1}$ ,可靠度函数  $R(t) = e^{-\lambda t}$ ,试求可靠度  $R = 99\%$  时的相应寿命  $t_{0.99}$ ,并求产品的中位寿命和特征寿命。

**解**

因为

$$R(t) = e^{-\lambda t}$$

故有

$$R(t_R) = R(t) = e^{-\lambda t_R}$$

两边取对数  $\ln R(t_R) = -\lambda t_R$

$$\text{得可靠度寿命 } t_{0.99} = -\frac{\ln R(t_{0.99})}{\lambda} = \frac{\ln(0.99)}{0.25 \times 10^{-4}} \text{h} = 402 \text{h}$$

$$\text{中位寿命 } t_{0.5} = -\frac{\ln R(t_{0.5})}{\lambda} = \frac{\ln(0.5)}{0.25 \times 10^{-4}} \text{h} = 27726 \text{h}$$

$$\text{特征寿命 } T_{0.1} = -\frac{\ln(e^{-1})}{\lambda} = \frac{\ln(0.3679)}{0.25 \times 10^{-4}} \text{h} = 40000 \text{h}$$

### 5.4 维修度

产品的维修性可用其维修度来衡量。维修度的定义是“对可能维修的产品在发生故障或失效后在规定的条件下和规定的时间  $(0, \tau)$  内完成修复的概率”,记为  $M(\tau)$ 。

与维修度相关的特征量还有平均修理时间和修复率。

平均修理时间 MTR (Mean Time To Repair) 是指可修复的产品的平均修理时间。

修复率  $\mu(\tau)$  是指“维修时间已达到某一时刻但尚未修复的产品在该时刻后的单位时间内完成修理的概率”。

### 5.5 有效度

有效度或称可用度,是指“可维修的产品在规定的条件下使用时,在某时刻  $t$  具有或维持其功能的概率”。有效度是综合可靠度与维修度的广义可靠性尺

度。

有效度  $A$  为工作时间对工作时间 (MTBF) 与维修时间 (MTTR) 之和的比,当工作时间和维修时间为指数分布时可表达为

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} = \frac{\mu}{\mu + \lambda}$$

式中  $\mu$  —— 失效率;

$\lambda$  —— 修复率。

### 5.6 可靠性特征量间的关系

可靠性特征量中  $R(t)$ 、 $F(t)$ 、 $f(t)$  和  $\lambda(t)$  是四个基本函数,只要知道其中一个,则所有特征量均可求得。基本函数间的关系见表 45.1-1。

表 45.1-1 可靠性特征量中四个基本函数关系

	$R(t)$	$F(t)$	$f(t)$	$\lambda(t)$
$R(t)$		$1 - F(t)$	$\int_t^{\infty} f(t) dt$	$e^{-\int_0^t \lambda(t) dt}$
$F(t)$	$1 - R(t)$		$\int_0^t f(t) dt$	$1 - e^{-\int_0^t \lambda(t) dt}$
$f(t)$	$-\frac{dR(t)}{dt}$	$\frac{dF(t)}{dt}$		$\lambda(t) e^{\int_0^t \lambda(t) dt}$
$\lambda(t)$	$-\frac{d}{dt} \ln R(t)$	$\frac{dF(t)}{dt} \frac{1}{1 - F(t)}$	$\frac{f(t)}{\int_t^{\infty} f(t) dt}$	

**例** 若失效率为常数  $\lambda$ ,求可靠度、累积失效概率、寿命概率密度、平均寿命、可靠寿命和中位寿命。

**解** 本例相当于偶然失效期,按表 45.1-1,可靠度

$$R(t) = e^{-\int_0^t \lambda dt} = e^{-\lambda t}$$

累积失效概率

$$F(t) = 1 - R(t) = 1 - e^{-\lambda t}$$

寿命概率密度

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t}$$

平均寿命

$$\theta = \int_0^{\infty} R(t) dt = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}$$

由可靠度函数解  $t$ ,可得可靠寿命

$$t(R) = \frac{1}{\lambda} \ln \frac{1}{R} = \theta \ln \frac{1}{R}$$

当取  $R = 0.5$ ,可得中位寿命

$$t = t(0.5) = \theta \ln \frac{1}{0.5} = 0.693 \theta$$

当可靠度较高时,例如  $R \geq 0.90$ ,将可靠度  $R = e^{-\lambda t}$  展开成级数,只取前两项,可得

$$R(t) \approx 1 - \lambda t = 1 - \frac{t}{\theta}$$

$$t(R) \approx (1 - R)\theta$$

本例是寿命为指数分布的情况时各可靠性特征量

的关系。一般复杂系统,寿命多服从这简单的指数分布,有时虽然并非指数分布,为了简化,应用平均失效率的概念,也可近似应用这些简单的关系。

### 6 可靠性设计程序和內容

可靠性设计一般程序和內容如图 45.1-19 所示。这个程序不仅要靠承担设计工作的人员,而且要靠从事质量管理、可靠性、生产工程、维修、服务、销售以及用户中技术人员的工作。在设计阶段不仅要使用传统设计所需的技术资料,而且必需参考质量管理、维修、使用、环境、市场等各种资料,收集处理这些情报并具体反映到设计中。

可靠性设计首先要明确产品的可靠性要求,确定可靠性目标。一般除特殊用户的产品外,很少给可靠性规格或目标值。通常是通过了解用户要求、竞争企业的动向、技术水平的现状和发展趋势等来确定。可靠性目标一般应包括功能、能源消耗、动力性能、可靠性指标以及安全性维修性等。

在确定产品可靠性目标值时,应该选择适合使用条件的可靠性特征量作为可靠性指标。各类产品常用的可靠性指标如表 45.1-2。表 45.1-3 列出了一些机械产品的主要可靠性指标,表 45.1-4 列出可靠度值分类

等级应用情况供设计参考。表 45.1-5 列出了不同零部件失效率参考值。

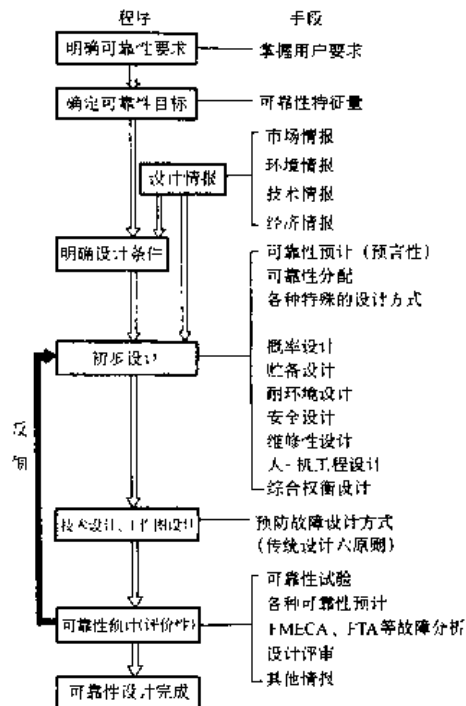


图 45.1-19 可靠性设计一般程序和內容

表 45.1-2 产品常用的可靠性指标

使用条件	连续使用				一次使用	
	可修复		不可修复		可修复	不可修复
可否修复	可修复		不可修复		可修复	不可修复
维修种类	预防维修	事后维修	用到耗损期	一定时间后报废	预防维修	
产品示例	电子系统、计算机、通信机、雷达、飞机、生产设备	家用电器、机械装置	电子元器件、机械零件、一般消费品	实际预防维修的零部件、广播设备用电子管	武器、过载继电器、救生器具	熔丝、闪光灯、雷管
常用指标	可靠度、有效度、平均无故障工作时间、平均修复时间	平均无故障工作时间、有效寿命、有效度	失效率、平均寿命	失效率、更换寿命	成功率	成功率

表 45.1-3 产品的主要可靠性指标值

产 品	指 标	可靠度	平均无故障工作限度	有效度	大修周期
载货汽车			10000km		
拖拉机			220~330h		
联合收割机			80~90h		
谷物联合收割机 1065 型			50~70h	0.90~0.95	
农用挂车			4500km		
推土机			400~500h	0.90~0.95	6000~10000h
单斗液压挖掘机			400~500h	0.90~0.95	8000~10000h
气腿式凿岩机 7655			1000m		
内燃凿岩机 YN30A			500m		
叉车			300~400h	0.90~0.95	6000~10000h
数控转塔车床 CK3263A(B)			200h	0.95	
卧式车床 CA6140			5000h	0.95	10a
精密丝杠车床 SG8630			4500h	0.95	7a
超高精度车床 SI-235			4500h	0.95	7a
无心磨床 M1040			120h	0.99	
高精度无心磨床 MG1050			100h	0.98	
通用桥式起重机				0.95	
汽车起重机			450h	0.80~0.90	8000~10000h
高速垂直制袋包装机 BZ-809			1000h		

(续)

产 品	指 标	可靠度	平均无故障 工作限度	有效度	大修周期
高速自动包装机 BZ-802			1000~1500h		6000~8000h
干草捡拾压捆机 C-D 型			10000~30000 捆		
拖拉机用发动机			1500~1700h		25000h
柴油机			1000~2000h		
小型汽油机(农用)			500~750h		
小型异步电机			20000h 以上		
活塞式制冷压缩机 8AS17			8000~12000h		
电站锅炉给水泵 50CHTA/6			10000h		
液压件中压叶片泵 YB 型			3000~10000h		
泥浆泵			3000~10000h		
灌注泵			5000~8000h		
渣浆泵			15000~25000h		
流程用泵			1000~5000h		
柱塞泵			4000h		
混流泵			4000h		
潜水电泵			4000h		
液压万能试验机 WE30.100		$R(1500h) \geq 0.90$			
洛氏硬度计 HR-150		$R(1000h) \geq 0.95$			
酸度计 PHS 3C, PHS 3D, PXD 38			3000h		
电度表 DJ862, DY862			10~20a		
水表			100000h		
家用电冰箱			20000h		
全自动洗衣机			1000h		
照像机			8000 次		
生物显微镜 44X, XS-B1, XSJ-1			4300h		
橡胶密封			3000h		
化工石油泵用机械密封			5000h		

表 45.1-4 可靠度分类等级应用举例

(续)

等级	可靠度	应 用 情 况
0	$< 0.9$	不重要的情况,失效后果可忽略不计,例如:不重要的轴承 $R=0.5 \sim 0.8$ ; 车辆低速齿轮 $R=0.8 \sim 0.9$
1	$\geq 0.9$	不很重要的情况,失效引起的损失不大,例如一般轴承 $R=0.90$ ,易维修的农机齿轮 $R \geq 0.90$ ,寿命长的汽轮机齿轮 $R \geq 0.98$
2	$\geq 0.99$	重要的情况,失效将引起大的损失,例如:一般齿轮的齿面强度 $R \approx 0.99$ ,弯强度 $R \approx 0.999$ ;高可靠性齿轮的齿面强度 $R \approx 0.999$ ,弯强度 $R \approx 0.9999$ ;寿命不长但要求高可靠性的飞机主传动齿轮 $R=0.99 \sim 0.9999$ 以上;高速轧机齿轮 $R=0.99 \sim 0.995$ ;建筑结构件,失效后果不严重的次要建筑 $R=0.997 \sim 0.9995$ (塑性破坏取低值,脆性破坏取高值,下同);失效后果严重的一般建筑 $R=0.9995 \sim 0.9999$ ;失效后果很严重的重要建筑 $R=0.9999 \sim 0.99999$
3	$\geq 0.999$	
4	$\geq 0.9999$	
5	1	很重要的情况,失效会引起灾难性后果,由于 $R > 0.9999$ ,其定量难以准确,建议在计算应力时取 $> 1$ 的计算系数来保证

表 45.1-5 不同零部件失效率参考值

零部件名称	$\lambda$ [失效数/(10 <sup>6</sup> h)]		
	最上限	平均	最下限
机床铸件(基础铸件)	0.7	0.175	0.015
一般轴承	1.0	0.5	0.02
球轴承(高速、重载)	3.53	1.8	0.075
球轴承(低速、轻载)	1.72	0.875	0.035
轴套或轴承	1.0	0.5	0.02
滚子轴承	1.0	0.5	0.02
凸轮	0.004	0.002	0.001
离合器	1.1	0.4	0.06
电磁离合器	0.93	0.6	0.45

零部件名称	$\lambda$ [失效数/(10 <sup>6</sup> h)]		
	最上限	平均	最下限
弹性联轴器	1.348	0.687	0.027
刚性联轴器	0.049	0.025	0.001
液压缸	0.12	0.008	0.005
气压缸	0.013	0.004	0.002
带传动	1.5	3.875	0.142
O形密封圈	0.08	0.02	0.01
橡胶密封圈	0.03	0.02	0.011
压力表	7.8	4.0	0.135
齿轮	0.20	0.12	0.0118
齿轮箱(运输用)	0.36	0.20	0.11
扇形齿轮	1.8	0.912	0.051
箱体	2.05	1.1	0.051
电动机	0.58	0.3	0.11
液压马达	7.15	4.3	1.45
转动密封	1.12	0.7	0.25
滑动密封	0.92	0.3	0.11
轴	0.62	0.35	0.15
弹簧	0.221	0.1125	0.004
弹簧(校准用)	0.42	0.22	0.009
弹簧(恢复力用)	0.022	0.012	0.001
阀门	8	5.1	2.0

试制新产品很少一次就成功。通常需经多次改进设计,逐步提高以达到可靠性目标值。在初步设计和技术设计告一段落,还需再进行可靠性预计,做必要的可靠性试验,对重要的部分用故障模式、效应及危害度分析(FMECA)、故障树分析(FTA)等方法进行可靠性、安全性分析,邀请有关各方面人员专家就可靠性进行评议审查。将设计的缺陷、潜在的故障原因、弥补的对策反馈给设计人员,进行改进设计,逐步完成可靠性设计。

## 第 2 章 可靠性设计的数学工具

概率论及数理统计是进行可靠性设计的数学基础。可靠性设计中将载荷、强度等设计变量作为随机变量,因而涉及到对随机变量的样本数据进行统计处理、判定分布类型,估计分布参数,计算可靠度等等。可靠性设计中常用的数理统计方法及工具简述如下。

### 1 可靠性设计常用的概率分布

载荷、强度、寿命等是可靠性设计涉及的重要的指标。这些指标一般都是随机变量,有一定的取值范围,服从一定的统计分布。

分布的类型很多,确定载荷、强度或产品的寿命服从何种分布的方法有两种,一种是根据其物理背景判断。产品的寿命分布与其所承受的载荷情况、产品的结构特点及其物理、化学、力学性能有关,与产品发生失效的物理过程有关。例如,通过失效分析,证实产品的故障模式或失效机理与某种类型分布的物理背景相接近时,可基本判定产品寿命服从该类型的分布。另一种方法是通过测试、试验及现场使用情况记录等获得数据,用统计推断的方法来判断它是属于何种分布。

在可靠性工程中,常用的分布有二项分布、泊松分布、指数分布、正态分布、对数正态分布和威布尔分布等。常用的分布列述如下。

#### 1.1 二项分布

设试验  $E$  只有两种可能的结果  $A$  和  $\bar{A}$ ,  $P(A) = p$ ,  $P(\bar{A}) = 1 - p$ 。用  $X$  表示在  $n$  重独立试验中事件  $A$  发生的次数,则  $X$  是一个随机变量,它的可能取值为  $0, 1, 2, \dots, k, \dots, n$ 。在这种情形  $X$  服从的概率分布称为二项分布,记为  $X \sim B(n, p)$ ,其概率分布为

$$P\{X = k\} = C_n^k p^k (1 - p)^{n-k} \quad (k = 0, 1, 2, \dots, n) \quad (45.2-1)$$

二项分布的数字特征为:  $E(X) = np$ ,  $D(X) = np(1 - p)$

二项分布的主要应用是在产品质量检验或可靠性抽样检验等问题中设计抽样检验方案。在可靠性设计中,二项分布可用于独立失效系统的可靠性计算和可靠性分配等。

#### 1.2 泊松(Poisson)分布

若随机变量  $X$  有如下的分布规律

$$P\{X = k\} = \frac{\mu^k e^{-\mu}}{k!} \quad (45.2-2)$$

则称其服从泊松分布,泊松分布的数字特征为

$$E(X) = \mu, D(X) = \mu$$

#### 1.3 正态(Gauss)分布

(1) 正态分布的概率密度函数为

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (-\infty < x < \infty) \quad (45.2-3)$$

记为  $X \sim N(\mu, \sigma^2)$ , 式中的  $\mu$  为均值,  $\sigma$  为标准差。正态分布是以  $x = \mu$  为对称轴的对称型分布,且在  $x = \mu$  处取得最大值,且改变  $\mu$  时,分布曲线发生平移,改变  $\sigma$  时对称轴位置不变。

(2) 标准正态分布

$\mu = 0, \sigma^2 = 1$  的正态分布称为标准正态分布,其概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (-\infty < x < \infty) \quad (45.2-4)$$

(3) 截尾正态分布

工程实际中有很多试验或观察数据近似服从正态分布。但正态分布的取值范围  $(-\infty$  到  $+\infty)$  不很符合实际情况。考虑到许多试验或观察数据无负值,因此用截尾正态分布来表示较为准确。截尾正态分布定义为:

若  $X$  是一个非负的随机变量,且密度函数为

$$f(x) = \frac{1}{\alpha \sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (0 \leq x < \infty) \quad (45.2-5)$$

则称  $X$  服从截尾正态分布。式中,  $\alpha$  为“正规化常数”,它保证了  $\int_0^{\infty} f(x) dx = 1$ 。

正态分布是最常用的分布,在机械可靠性设计中,可以用来描述零件的强度分布。从物理背景上讲,如果影响某个随机变量的独立因素很多,且不存在起决定作用的主导因素时,则该随机变量一般可用正态分布来描述。正态分布随机变量的取值范围从负无穷大直到正无穷大,从这一点来看,强度不可能是真正的正态分布,而只可能是截尾正态分布。

#### 1.4 对数正态分布

若  $X$  是一个随机变量,且  $Y = \ln(X)$  服从正态分

布,即

$$Y = \ln(X) \sim N(\mu, \sigma^2)$$

则称  $X$  服从对数正态分布,其概率密度函数为

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] \quad (45.2-6)$$

对数正态分布的均值与方差分别为  $E(X) = e^{\mu + \frac{\sigma^2}{2}}$  和  $D(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$ 。

对数正态分布是一种偏态分布,是描述零件疲劳寿命分布的一种主要分布形式。例如,对数正态分布可以用来描述螺旋弹簧、螺栓、齿轮、以及轴类零件的疲劳寿命分布。对数正态分布使用方便,能给出可靠的置信区间,从而在寿命试验中得到了广泛的应用。

### 1.5 威布尔(Weibull)分布

#### (1) 三参数威布尔分布

三参数威布尔分布记为  $X \sim W(k, b, a, x)$ , 其中  $k$  为形状参数,  $b$  为尺度参数,  $a$  为位置参数。三参数威布尔分布的概率密度函数为

$$f(x) = \frac{k}{b} \left(\frac{x-a}{b}\right)^{k-1} \exp\left[-\left(\frac{x-a}{b}\right)^k\right] \quad x \geq a, m > 0, b > 0 \quad (45.2-7)$$

三参数威布尔分布的分布函数为

$$F(x) = P\{X \leq x\} = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^k\right] \quad (45.2-8)$$

三参数威布尔分布的数字特征为  $E(X) = a + b\Gamma\left(1 + \frac{1}{k}\right)$ ,  $D(X) = b^2 \left\{ \Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right\}$ 。

#### (2) 两参数威布尔分布

或把三参数威布尔分布的位置参数  $a$  取为 0, 则简化为两参数威布尔分布。其密度函数为

$$f(x) = \frac{k}{b} \left(\frac{x}{b}\right)^{k-1} \exp\left[-\left(\frac{x}{b}\right)^k\right] \quad x \geq 0, k > 0, b > 0 \quad (45.2-9)$$

两参数威布尔分布的分布函数为

$$F(x) = P\{X \leq x\} = 1 - \exp\left[-\left(\frac{x}{b}\right)^k\right] \quad (45.2-10)$$

两参数威布尔分布的数字特征为  $E(X) = b\Gamma\left(1 + \frac{1}{k}\right)$ ,  $D(X) = b^2 \left\{ \Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right\}$ 。

瑞典工程师威布尔(Weibull)采用“链式”模型来研究、描述结构强度和寿命问题,假设一个结构是由若干小元件(设为  $n$  个)串联而成,将结构看成是由  $n$  个环构成的一条链条,其强度(或寿命)取决于最薄弱环节的强度(或寿命)。单个链的强度(或寿命)为一随机变

量。设各环强度(或寿命)相互独立,分布相同,则求链强度(或寿命)的概率分布就变成求极小值分布问题,由此得出了威布尔分布函数。

由于威布尔分布是根据最薄弱环节模型或串联模型得到的,能充分反映材料缺陷和应力集中源对材料疲劳寿命的影响,所以,将它作为材料或零件的寿命分布模型或给定寿命下的疲劳强度模型是比较合适的。

### 1.6 指数分布

指数分布的概率密度函数为

$$f(x) = \lambda e^{-\lambda x} \quad (x \geq 0, \lambda > 0) \quad (45.2-11)$$

式中  $\lambda$  为常数,是指数分布的失效率。

指数分布的分布函数为

$$F(x) = 1 - e^{-\lambda x} \quad (45.2-12)$$

若令  $\theta = 1/\lambda$  为指数分布的平均寿命,用  $t$  表示失效时间随机变量,指数分布的概率密度函数可表达为

$$f(t) = \frac{1}{\theta} e^{-t/\theta} \quad (45.2-13)$$

$$F(t) = 1 - e^{-t/\theta} \quad (45.2-14)$$

指数分布的数字特征为  $E(x) = 1/\lambda$  (或  $E(t) = \theta$ ),  $D(x) = 1/\lambda^2$  (或  $D(t) = \theta^2$ )。

指数分布的可靠度函数为

$$R(t) = e^{-t/\theta} = e^{-\lambda t} \quad (45.2-15)$$

上式与泊松分布展开式的第一项完全相同,这说明若产品在一定时间区间内的失效数服从泊松分布,则该产品的寿命服从指数分布。

指数分布的失效率为

$$\lambda(t) = 1/\theta = \lambda \quad (45.2-16)$$

指数分布的一个重要性质是无记忆性,可表达为

$$P\{T > t_0 + t \mid \{T > t_0\}\} = P\{T > t_0\} = e^{-\lambda t} \quad (45.2-17)$$

指数分布可以看作当威布尔分布中的形状参数等于 1 时的特殊情形。

在电子元器件的可靠性研究中,指数分布应用广泛,在日本的工业标准和美国军用标准中,半导体器件的抽验方案都是采用指数分布。此外,指数分布还可用来描述大型复杂系统的故障间隔时间的分布,特别是在部件或机器的整机试验中得到了广泛的应用。

### 1.7 I 型极值分布(Gumbel 分布)

(1) I 型极大值分布概率密度函数为:

$$f(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-e^{-\frac{x-\mu}{\sigma}}}$$

$$(-\infty < x < \infty, \sigma > 0, -\infty < \mu < \infty)$$

$$(45.2-18)$$

式中  $\mu$  为极值分布的位置参数;



$\sigma$ ——极值分布的尺度参数。

极大值分布的数字特征为： $E(x) = \mu + 0.5772\sigma$ ，

$$D(x) = 1.644\sigma^2$$

(2) I 型极小值分布

I 型极小值分布概率密度函数为

$$f(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-\frac{x-\mu}{\sigma}}$$

$$(-\infty < x < \infty, 0 < \sigma < \infty, -\infty < \mu < \infty)$$

(45.2-19)

式中  $\mu$ ——位置参数；

$\sigma$ ——尺度参数。

极小值分布的数字特征为： $E(x) = \mu + 0.5772\sigma$ ，

$$D(x) = 1.644\sigma^2$$

I 型极值分布是对应于大量子样的最小值或最大值的分布，主要用来描述一个随机变量出现极小值或极大值的现象。这类问题很多，例如，建筑工程中结构构件抗力的最小值分布，结构载荷的最大值分布；机械工程中导致机械产品失效的强度或寿命的最小值分布，短期过载的最大值分布，串联系统的最弱元件，并联系统中最强元件等等。

I 型极小值分布与威布尔分布的关系为：设随机变量  $T$  服从威布尔分布，则  $X = \ln T$  服从 I 型最小极值分布。由于 I 型最小极值分布与威布尔分布有这种关系，因此，可以用它作为威布尔分布的拟合优度检验。所以，极值分布也是寿命数据分析中经常用到的一种分布类型。

### 1.8 $\Gamma$ 分布、瑞利分布、 $\beta$ 分布

(1)  $\Gamma$  分布(伽玛分布)

$\Gamma$  分布(伽玛分布)概率密度函数为

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (45.2-20)$$

式中  $\beta$ ——形状参数， $\beta > 0$ ；

$\alpha$ ——尺度参数， $\alpha > 0$ 。

$\Gamma$  分布的数字特征为： $E(X) = \frac{\alpha}{\beta}$ ， $D(X) = \frac{\alpha}{\beta^2}$ 。

$\Gamma$  分布适用性较广，但参数估计复杂，不如威布尔分布使用方便。

(2) 瑞利分布

瑞利分布概率密度函数为

$$f(x) = \begin{cases} \frac{x}{\mu^2} e^{-\frac{x^2}{2\mu^2}} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (45.2-21)$$

式中  $\mu$ ——为常数， $\mu > 0$ ；

瑞利分布的数字特征为： $E(X) = \sqrt{\frac{\pi}{2}} \mu$ ， $D(X) =$

$$\frac{(1-\pi)\mu^2}{2}$$

(3)  $\beta$  分布(贝塔分布)

$\beta$  分布概率密度函数为

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)}, & 0 < x < 1 \\ 0 & x \leq 0 \text{ 或 } x \geq 1 \end{cases}$$

(45.2-22)

式中  $\alpha, \beta$ ——常数， $\beta > 0, \alpha > 0$ ；

$\beta$  分布的数字特征为： $E(X) = \frac{\alpha}{\alpha+\beta}$ ， $D(X) =$

$$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

### 1.9 $\chi^2$ 分布、 $t$ 分布、 $F$ 分布

(1)  $\chi^2$  分布

设  $X$  为总体， $X_1, X_2, \dots, X_n$  为子样， $x_1, x_2, \dots, x_n$  为观测值。若总体  $X \sim N(0, 1)$ ， $X_1, X_2, \dots, X_n$  是总体的一个样本，统计量  $g = \sum_{i=1}^n x_i^2$ ，则统计量  $g$  服从自由度为  $n$  的  $\chi^2$  分布，记为  $\chi^2 \sim \chi^2(n)$ 。

$\chi^2$  分布的概率密度函数为

$$f(x) = \begin{cases} \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(45.2-23)

$\chi^2$  分布的数字特征： $E(\chi^2) = n$ ， $D(\chi^2) = 2n$

$\chi^2$  分布常用来解决有关母体标准差的统计推断和母体分布的假设检验问题。

(2)  $t$  分布

若总体  $X \sim N(0, 1)$ ， $Y \sim \chi^2(n)$ ，且  $X$  与  $Y$  相互独立，则称统计量  $T = \frac{X}{\sqrt{Y/n}}$  服从自由度为  $n$  的  $t$  分布，

记为  $T \sim t(n)$ ，其分布函数为

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

$-\infty < t < +\infty$  (45.2-24)

$t$  分布的数字特征： $E(\chi^2) = 0$ ， $D(\chi^2) = \frac{n}{n-2}$ 。

$t$  分布是用于对正态分布的母体的均值进行估计的一个应用十分广泛的分布密度函数。 $t$  分布密度曲线与标准正态密度分布曲线相类似。从理论上可以证明：当自由度  $n \rightarrow \infty$  时， $t$  分布趋于标准正态分布。实际上，当  $n \geq 30$  时，两者已十分相近。用  $t$  分布可检验一个小子样是否来自已知平均值的母体，或作正态母体

平均值的区间估计。

(3)  $F$  分布

设  $\chi_1^2$  和  $\chi_2^2$  为两个相互独立的变量, 并且已知  $\chi_1^2$  是自由度为  $n_1$  的  $\chi^2$  概率密度函数,  $\chi_2^2$  是自由度为  $n_2$  的  $\chi^2$  概率密度函数, 它们的定义域皆为  $(0, \infty)$ , 随机变量  $F_r$  定义为  $\chi_1^2$  和  $\chi_2^2$  比值:  $F_r = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$

则统计量  $F_r$  服从  $F$  分布, 记为  $F \sim F(n_1, n_2)$ , 其概率密度函数为

$$f_F(x) = \begin{cases} \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} \Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{x^{\frac{n_1}{2}-1}}{(n_2+n_1x)^{\frac{n_1+n_2}{2}}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(45.2.25)

表 45.2-1 可靠性分析常用的概率分布

名称记号	概率分布及其定义域、参数条件	均值 $E(X)$	方差 $D(X)$	图 形
二项分布 $B(n, p)$	$P_B(X=x) = \binom{n}{x} p^x q^{n-x}$ $x=0, 1, \dots, n$ $0 < p < 1$ , $p+q=1$ $n$ 为正整数	$np$	$npq$	
泊松分布 $P(\lambda)$	$P_P(X=x) = \frac{\lambda^x}{x!} e^{-\lambda}$ $x=0, 1, 2, \dots$ $\lambda > 0$	$\lambda$	$\lambda$	
超几何分布 $H(n, M, N)$	$P_H(X=x) = \frac{\binom{N-M}{n-x} \binom{M}{x}}{\binom{N}{n}}$ $x = \max(0, n-N+M), \dots, \min(n, M)$ $N, M, n$ 为正整数 $0 \leq M \leq N, 0 \leq n \leq N$	$\frac{nM}{N}$	$\frac{n(N-n)(N-M)M}{N^2(N-1)}$	
均匀分布 $u(a, b)$	$f_u(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ 或 } x > b \end{cases}$ $-\infty < a < b < \infty$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
正态分布 (高斯分布) $N(\mu, \sigma^2)$	$f_N(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$ $-\infty < \mu < \infty, \sigma > 0$	$\mu$	$\sigma^2$	

$F$  分布的数字特征

$$E(\chi^2) = \frac{n_2}{n_2 - 2}, \quad n_2 > 2$$

$$D(\chi^2) = \frac{2n_2^2(n_2 + n_2 - 2)}{n_2(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4$$

用  $F$  分布进行的检验称为  $F$  检验, 可用来检验两个小子样是否来自标准差相同的两个母体。

可靠性分析中常用的概率表达式、图形及其特征量如表 45.2-1 所示。表 45.2-2 为  $\Gamma$  分布表, 表 45.2-3 为二项分布表, 表 45.2-4 为标准正态分布表, 表 45.2-5 为  $\chi^2$  分布表, 表 45.2-6 为  $t$  分布表, 表 45.2.7 为  $F$  分布表。

(续)

名称记号	概率分布及其定义域、参数条件	均值 $E(X)$	方差 $D(X)$	图 形
对数正态分布(自然对数) $\ln(\mu, \sigma^2)$ 对数正态分布(常用对数) $\lg(\mu, \sigma^2)$	$f_{\ln}(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & x > 0 \\ 0, & x \leq 0 \end{cases}$ $f_{\lg}(x) = \begin{cases} \frac{\lg e}{\sigma \lambda \sqrt{2\pi}} e^{-\frac{(\lg x - \mu)^2}{2\sigma^2}} & x > 0 \\ 0, & x \leq 0 \end{cases}$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^2 \mu - \sigma^2 (e^{\sigma^2} - 1)$	
威布尔分布(Ⅲ型极值分布) $W(k, a, b)$	$f_W(x) = \begin{cases} \frac{k}{b} \left(\frac{x-a}{b}\right)^{k-1} \times e^{-\left(\frac{x-a}{b}\right)^k} & x \geq a \\ 0, & x < a \end{cases}$ <p>形状参数 <math>k &gt; 0</math> 尺度参数 <math>b &gt; 0</math> 位置参数 <math>a</math></p>	$b\Gamma\left(1 + \frac{1}{k}\right) + a$	$b^2 \left[ \Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right]$	
指数分布 $e(\lambda)$	$f_e(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ <p><math>\lambda &gt; 0</math></p>	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	
瑞利分布 $R(\mu)$	$f_R(x) = \begin{cases} \frac{x}{\mu^2} e^{-\frac{x^2}{2\mu^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ <p><math>\mu &gt; 0</math></p>	$\sqrt{\frac{\pi}{2}} \mu$	$\frac{(1-\pi)\mu^2}{2}$	
$\beta$ 分布(贝塔分布) $\beta(a, \beta)$ 贝塔函数 $\beta(a, \beta) = \frac{\Gamma(a)\Gamma(\beta)}{\Gamma(a+\beta)}$ $a > 0, \beta > 0$	$f_\beta(x) = \begin{cases} \frac{x^{a-1}(1-x)^{\beta-1}}{\beta(a, \beta)}, & 0 < x < 1 \\ 0, & x \leq 0 \text{ 或 } x \geq 1 \end{cases}$	$\frac{a}{a+\beta}$	$\frac{a\beta}{(a+\beta-1)(a+\beta)^2}$	
$\Gamma$ 分布(伽马分布) $\Gamma(a, \beta)$	$f_\Gamma(x) = \begin{cases} \frac{\beta^a}{\Gamma(a)} x^{a-1} e^{-\beta x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{a}{\beta}$	$\frac{a}{\beta^2}$	
最大Ⅰ型极值分布(贡贝尔分布) $M(\sigma, \mu)$	$f_M(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-\frac{x-\mu}{\sigma}}$ <p><math>-\infty &lt; x &lt; \infty</math> <math>-\infty &lt; \mu &lt; \infty</math> <math>0 &lt; \sigma</math></p>	$\mu + \gamma\sigma$ 欧拉常数 $\gamma \approx 0.5772$	$\frac{\pi^2}{6} \sigma^2 \approx (1.283\sigma)^2$	
最小Ⅰ型极值分布(贡贝尔分布) $m(\sigma, \mu)$	$f_m(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-\frac{x-\mu}{\sigma}}$ <p><math>-\infty &lt; x &lt; \infty</math> <math>-\infty &lt; \mu &lt; \infty</math> <math>0 &lt; \sigma</math></p>	$\mu - \gamma\sigma$ 欧拉常数 $\gamma \approx 0.5772$	$\frac{\pi^2}{6} \sigma^2 \approx (1.283\sigma)^2$	

(续)

名称记号	概率分布及其定义域、参数条件	均值 $E(X)$	方差 $D(X)$	图 形
$\chi^2$ 分布 $\chi^2(\nu)$	$f_x^2(x) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \times e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$ 自由度 $\nu$ 为正整数	$\nu$	$2\nu$	
$t$ 分布 (学生分布) $t(\nu)$	$f_t(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu} \Gamma(\frac{\nu}{2})} \times \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ 自由度 $\nu$ 为正整数	0 ( $\nu > 1$ )	$\frac{\nu}{\nu-2}$ ( $\nu > 2$ )	
$F$ 分布 $F(\nu_1, \nu_2)$	$f_F(x) = \begin{cases} \frac{x^{\frac{\nu_1}{2}-1} (1-x)^{\frac{\nu_2}{2}-1}}{\beta(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \times \left(1 + \frac{\nu_1 x}{\nu_2}\right)^{-\frac{\nu_1+\nu_2}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 贝塔函数 $\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) = \frac{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2})}{\Gamma(\frac{\nu_1}{2} + \frac{\nu_2}{2})}$ 自由度 $\nu_1, \nu_2$ 为正整数	$\frac{\nu_2}{\nu_2-2}$ ( $\nu_2 > 2$ )	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$ ( $\nu_2 > 4$ )	

表 45.2-2  $\Gamma$  函数表

$x$	$\Gamma(x)$									
	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
1.00	1.0000	0.9994	0.9988	0.9983	0.9977	0.9971	0.9966	0.9960	0.9954	0.9949
1.01	0.9943	0.9938	0.9932	0.9927	0.9921	0.9916	0.9910	0.9905	0.9899	0.9894
1.02	0.9888	0.9883	0.9878	0.9872	0.9867	0.9862	0.9856	0.9851	0.9846	0.9841
1.03	0.9835	0.9830	0.9825	0.9820	0.9815	0.9810	0.9805	0.9800	0.9794	0.9789
1.04	0.9784	0.9779	0.9774	0.9769	0.9764	0.9759	0.9755	0.9750	0.9745	0.9740
1.05	0.9735	0.9730	0.9725	0.9721	0.9716	0.9711	0.9706	0.9702	0.9697	0.9692
1.06	0.9687	0.9683	0.9678	0.9673	0.9669	0.9664	0.9660	0.9655	0.9651	0.9646
1.07	0.9642	0.9637	0.9633	0.9628	0.9624	0.9619	0.9615	0.9610	0.9606	0.9602
1.08	0.9597	0.9593	0.9589	0.9584	0.9580	0.9576	0.9571	0.9567	0.9563	0.9559
1.09	0.9555	0.9550	0.9546	0.9542	0.9538	0.9534	0.9530	0.9526	0.9522	0.9518
1.10	0.9514	0.9509	0.9505	0.9501	0.9498	0.9494	0.9490	0.9486	0.9482	0.9478
1.11	0.9474	0.9470	0.9466	0.9462	0.9459	0.9455	0.9451	0.9447	0.9443	0.9440
1.12	0.9436	0.9432	0.9428	0.9425	0.9421	0.9417	0.9414	0.9410	0.9407	0.9403
1.13	0.9399	0.9396	0.9392	0.9389	0.9385	0.9382	0.9378	0.9375	0.9371	0.9368
1.14	0.9364	0.9361	0.9357	0.9354	0.9350	0.9347	0.9344	0.9340	0.9337	0.9334
1.15	0.9330	0.9327	0.9324	0.9321	0.9317	0.9314	0.9311	0.9308	0.9304	0.9301
1.16	0.9298	0.9295	0.9292	0.9289	0.9285	0.9282	0.9279	0.9276	0.9273	0.9270
1.17	0.9267	0.9264	0.9261	0.9258	0.9255	0.9252	0.9249	0.9246	0.9243	0.9240
1.18	0.9237	0.9234	0.9231	0.9229	0.9226	0.9223	0.9220	0.9217	0.9214	0.9212
1.19	0.9209	0.9206	0.9203	0.9201	0.9198	0.9195	0.9192	0.9190	0.9187	0.9184
1.20	0.9182	0.9179	0.9176	0.9174	0.9171	0.9169	0.9166	0.9163	0.9161	0.9158
1.21	0.9156	0.9153	0.9151	0.9148	0.9146	0.9143	0.9141	0.9138	0.9136	0.9133
1.22	0.9131	0.9129	0.9126	0.9124	0.9122	0.9119	0.9117	0.9114	0.9112	0.9110
1.23	0.9108	0.9105	0.9103	0.9101	0.9098	0.9096	0.9094	0.9092	0.9090	0.9087
1.24	0.9085	0.9083	0.9081	0.9079	0.9077	0.9074	0.9072	0.9070	0.9068	0.9066
1.25	0.9064	0.9062	0.9060	0.9058	0.9056	0.9054	0.9052	0.9050	0.9048	0.9046
1.26	0.9044	0.9042	0.9040	0.9038	0.9036	0.9034	0.9032	0.9031	0.9029	0.9027

(续)

x	Γ(x)									
	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
1.27	0.9025	0.9023	0.9021	0.9020	0.9018	0.9016	0.9014	0.9012	0.9011	0.9009
1.28	0.9007	0.9005	0.9004	0.9002	0.9000	0.8999	0.8997	0.8995	0.8994	0.8992
1.29	0.8990	0.8989	0.8987	0.8986	0.8984	0.8982	0.8981	0.8979	0.8978	0.8976
1.30	0.8975	0.8973	0.8972	0.8970	0.8969	0.8967	0.8966	0.8964	0.8963	0.8961
1.31	0.8960	0.8959	0.8957	0.8956	0.8954	0.8953	0.8952	0.8950	0.8949	0.8948
1.32	0.8946	0.8945	0.8944	0.8943	0.8941	0.8940	0.8939	0.8937	0.8936	0.8935
1.33	0.8934	0.8933	0.8931	0.8930	0.8929	0.8928	0.8927	0.8926	0.8924	0.8923
1.34	0.8922	0.8921	0.8920	0.8919	0.8918	0.8917	0.8916	0.8915	0.8914	0.8913
1.35	0.8912	0.8911	0.8910	0.8909	0.8908	0.8907	0.8906	0.8905	0.8904	0.8903
1.36	0.8902	0.8901	0.8900	0.8899	0.8898	0.8897	0.8897	0.8896	0.8895	0.8894
1.37	0.8893	0.8892	0.8892	0.8891	0.8890	0.8889	0.8888	0.8888	0.8887	0.8886
1.38	0.8885	0.8885	0.8884	0.8883	0.8883	0.8882	0.8881	0.8880	0.8880	0.8879
1.39	0.8879	0.8878	0.8877	0.8877	0.8876	0.8875	0.8875	0.8874	0.8874	0.8873
1.40	0.8873	0.8872	0.8872	0.8871	0.8871	0.8870	0.8870	0.8869	0.8869	0.8868
1.41	0.8868	0.8867	0.8867	0.8866	0.8866	0.8865	0.8865	0.8865	0.8864	0.8864
1.42	0.8864	0.8863	0.8863	0.8863	0.8862	0.8862	0.8862	0.8861	0.8861	0.8861
1.43	0.8860	0.8860	0.8860	0.8860	0.8859	0.8859	0.8859	0.8859	0.8858	0.8858
1.44	0.8858	0.8858	0.8858	0.8858	0.8857	0.8857	0.8857	0.8857	0.8857	0.8857
1.45	0.8857	0.8857	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856
1.46	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856	0.8856
1.47	0.8856	0.8856	0.8856	0.8857	0.8857	0.8857	0.8857	0.8857	0.8857	0.8857
1.48	0.8857	0.8858	0.8858	0.8858	0.8858	0.8858	0.8858	0.8859	0.8859	0.8859
1.49	0.8859	0.8860	0.8860	0.8860	0.8860	0.8861	0.8861	0.8861	0.8862	0.8862
1.50	0.8862	0.8863	0.8863	0.8863	0.8864	0.8864	0.8864	0.8865	0.8865	0.8866
1.51	0.8866	0.8866	0.8867	0.8867	0.8868	0.8868	0.8869	0.8869	0.8869	0.8870
1.52	0.8870	0.8871	0.8871	0.8872	0.8872	0.8873	0.8873	0.8874	0.8875	0.8875
1.53	0.8876	0.8876	0.8877	0.8877	0.8878	0.8879	0.8879	0.8880	0.8880	0.8881
1.54	0.8882	0.8882	0.8883	0.8884	0.8884	0.8885	0.8886	0.8887	0.8887	0.8888
1.55	0.8889	0.8889	0.8890	0.8891	0.8892	0.8892	0.8893	0.8894	0.8895	0.8896
1.56	0.8896	0.8897	0.8898	0.8899	0.8900	0.8901	0.8901	0.8902	0.8903	0.8904
1.57	0.8905	0.8906	0.8907	0.8908	0.8909	0.8909	0.8910	0.8911	0.8912	0.8913
1.58	0.8914	0.8915	0.8916	0.8917	0.8918	0.8919	0.8920	0.8921	0.8922	0.8923
1.59	0.8924	0.8925	0.8926	0.8927	0.8929	0.8930	0.8931	0.8932	0.8933	0.8934
1.60	0.8935	0.8936	0.8937	0.8939	0.8940	0.8941	0.8942	0.8943	0.8944	0.8946
1.61	0.8947	0.8948	0.8949	0.8950	0.8952	0.8953	0.8954	0.8955	0.8957	0.8958
1.62	0.8959	0.8961	0.8962	0.8963	0.8964	0.8966	0.8967	0.8968	0.8970	0.8971
1.63	0.8972	0.8974	0.8975	0.8977	0.8978	0.8979	0.8981	0.8982	0.8984	0.8985
1.64	0.8986	0.8988	0.8989	0.8991	0.8992	0.8994	0.8995	0.8997	0.8998	0.9000
1.65	0.9001	0.9003	0.9004	0.9006	0.9007	0.9009	0.9010	0.9012	0.9014	0.9015
1.66	0.9017	0.9018	0.9020	0.9021	0.9023	0.9025	0.9026	0.9028	0.9030	0.9031
1.67	0.9033	0.9035	0.9036	0.9038	0.9040	0.9041	0.9043	0.9045	0.9047	0.9048
1.68	0.9050	0.9052	0.9054	0.9055	0.9057	0.9059	0.9061	0.9062	0.9064	0.9066
1.69	0.9068	0.9070	0.9071	0.9073	0.9075	0.9077	0.9079	0.9081	0.9083	0.9084
1.70	0.9086	0.9088	0.9090	0.9092	0.9094	0.9096	0.9098	0.9100	0.9102	0.9104
1.71	0.9106	0.9108	0.9110	0.9112	0.9114	0.9116	0.9118	0.9120	0.9122	0.9124
1.72	0.9126	0.9128	0.9130	0.9132	0.9134	0.9136	0.9138	0.9140	0.9142	0.9145
1.73	0.9147	0.9149	0.9151	0.9153	0.9155	0.9157	0.9160	0.9162	0.9164	0.9166
1.74	0.9168	0.9170	0.9173	0.9175	0.9177	0.9179	0.9182	0.9184	0.9186	0.9188
1.75	0.9191	0.9193	0.9195	0.9197	0.9200	0.9202	0.9204	0.9207	0.9210	0.9211
1.76	0.9214	0.9216	0.9218	0.9221	0.9223	0.9226	0.9228	0.9230	0.9233	0.9235
1.77	0.9238	0.9240	0.9242	0.9245	0.9247	0.9250	0.9252	0.9255	0.9257	0.9260
1.78	0.9262	0.9265	0.9267	0.9270	0.9272	0.9275	0.9277	0.9280	0.9283	0.9285
1.79	0.9288	0.9290	0.9293	0.9295	0.9298	0.9301	0.9303	0.9306	0.9309	0.9311
1.80	0.9314	0.9316	0.9319	0.9322	0.9325	0.9327	0.9330	0.9333	0.9335	0.9338
1.81	0.9341	0.9343	0.9346	0.9349	0.9352	0.9355	0.9357	0.9360	0.9363	0.9366
1.82	0.9368	0.9371	0.9374	0.9377	0.9380	0.9383	0.9385	0.9388	0.9391	0.9394
1.83	0.9397	0.9400	0.9403	0.9406	0.9408	0.9411	0.9414	0.9417	0.9420	0.9423
1.84	0.9426	0.9429	0.9432	0.9435	0.9438	0.9441	0.9444	0.9447	0.9450	0.9453
1.85	0.9456	0.9459	0.9462	0.9465	0.9468	0.9471	0.9474	0.9478	0.9481	0.9484
1.86	0.9487	0.9490	0.9493	0.9496	0.9499	0.9503	0.9506	0.9509	0.9512	0.9515
1.87	0.9518	0.9522	0.9525	0.9528	0.9531	0.9534	0.9538	0.9541	0.9544	0.9547
1.88	0.9551	0.9554	0.9557	0.9561	0.9564	0.9567	0.9570	0.9574	0.9577	0.9580
1.89	0.9584	0.9587	0.9591	0.9594	0.9597	0.9601	0.9604	0.9607	0.9611	0.9614
1.90	0.9618	0.9621	0.9625	0.9628	0.9631	0.9635	0.9638	0.9642	0.9645	0.9649
1.91	0.9652	0.9656	0.9659	0.9663	0.9666	0.9670	0.9673	0.9677	0.9681	0.9684
1.92	0.9688	0.9691	0.9695	0.9699	0.9702	0.9706	0.9709	0.9713	0.9717	0.9720
1.93	0.9724	0.9728	0.9731	0.9735	0.9739	0.9742	0.9746	0.9750	0.9754	0.9757
1.94	0.9761	0.9765	0.9768	0.9772	0.9776	0.9780	0.9784	0.9787	0.9791	0.9795
1.95	0.9799	0.9803	0.9806	0.9810	0.9814	0.9818	0.9822	0.9826	0.9830	0.9834
1.96	0.9837	0.9841	0.9845	0.9849	0.9853	0.9857	0.9861	0.9865	0.9869	0.9873
1.97	0.9877	0.9881	0.9885	0.9889	0.9893	0.9897	0.9901	0.9905	0.9909	0.9913
1.98	0.9917	0.9921	0.9925	0.9929	0.9933	0.9938	0.9942	0.9946	0.9950	0.9954
1.99	0.9958	0.9962	0.9966	0.9971	0.9975	0.9979	0.9983	0.9987	0.9992	0.9996

对  $x < 1$  或  $x > 2$  的伽马函数值, 可以利用下式算出:  $\Gamma(x) = \frac{\Gamma(x+1)}{x}$ ,  $\Gamma(x) = (x-1)\Gamma(x-1)$

例: (1)  $\Gamma(0.8) = \frac{\Gamma(1.8)}{0.8} = \frac{0.9314}{0.8} = 1.164$  (2)  $\Gamma(2.5) = 1.5 \times \Gamma(1.5) = 1.5 \times 0.8862 = 1.329$

表 45.2-3 二项分布表

$$\sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}$$

n	k	p									p	k	n
		0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.3	0.4			
5	5			0.00000	0.00000	0.00000	0.00001	0.00032	0.00243	0.01024	0.03125	5	5
	4	0.00000	0.00000	0.00001	0.00006	0.00019	0.00046	0.00672	0.03078	0.08704	0.18750	4	
	3	0.00001	0.00008	0.00060	0.00197	0.00453	0.00856	0.05792	0.16308	0.31744	0.50000	3	
	2	0.00098	0.00384	0.01476	0.03187	0.05436	0.08146	0.26272	0.47178	0.66304	0.81250	2	
	1	0.04901	0.09608	0.18463	0.26610	0.34092	0.40951	0.67232	0.83193	0.92224	0.96875	1	
10	10								0.00001	0.00010	0.00098	10	10
	9							0.00000	0.00014	0.00168	0.01074	9	
	8						0.00000	0.00008	0.00159	0.01229	0.05469	8	
	7				0.00000	0.00000	0.00001	0.00086	0.01059	0.03476	0.17483	7	
	6			0.00000	0.00001	0.00004	0.00015	0.00637	0.04735	0.16624	0.37695	6	
	5		0.00000	0.00002	0.00013	0.00059	0.00163	0.03279	0.15027	0.36690	0.62305	5	
	4	0.00000	0.00003	0.00044	0.00203	0.00580	0.01280	0.12087	0.35039	0.61772	0.82813	4	
	3	0.00011	0.00086	0.00621	0.01884	0.04008	0.07019	0.32220	0.61722	0.83271	0.94531	3	
	2	0.00427	0.01618	0.05815	0.11759	0.18788	0.26390	0.62419	0.85069	0.95364	0.98926	2	
	1	0.09562	0.18293	0.33517	0.46133	0.56561	0.65132	0.89263	0.97175	0.99395	0.99902	1	
15	15									0.00000	0.00003	15	15
	14								0.00000	0.00003	0.00049	14	
	13								0.00001	0.00023	0.00369	13	
	12							0.00000	0.00009	0.00193	0.01758	12	
	11							0.00001	0.00067	0.00935	0.05923	11	
	10							0.00011	0.00365	0.03383	0.15088	10	
	9					0.00000	0.00000	0.00079	0.01524	0.09505	0.30362	9	
	8				0.00000	0.00001	0.00003	0.00424	0.05001	0.21310	0.50000	8	
	7			0.00000	0.00001	0.00006	0.00031	0.01806	0.13114	0.39019	0.69638	7	
	6		0.00000	0.00001	0.00015	0.00070	0.00225	0.06105	0.27833	0.59678	0.84912	6	
	5	0.00000	0.00001	0.00022	0.00140	0.00497	0.01272	0.16423	0.48451	0.78272	0.94077	5	
	4	0.00001	0.00018	0.00245	0.01036	0.02731	0.05556	0.35184	0.70313	0.90950	0.98242	4	
	3	0.00042	0.00304	0.02029	0.05713	0.11297	0.18406	0.60198	0.87317	0.97289	0.99631	3	
	2	0.00963	0.03534	0.11911	0.22624	0.31027	0.45096	0.83287	0.96473	0.99483	0.99951	2	
	1	0.13994	0.26143	0.45791	0.60471	0.71370	0.79411	0.96482	0.99525	0.99953	0.99997	1	
20	20										0.00000	20	20
	19									0.00000	0.00002	19	
	18									0.00001	0.00020	18	
	17								0.00000	0.00005	0.00120	17	
	16								0.00001	0.00032	0.00591	16	
	15								0.00004	0.00161	0.02069	15	
	14							0.00000	0.00026	0.00647	0.06766	14	
	13							0.00002	0.00128	0.02103	0.13159	13	
	12							0.00010	0.00514	0.05658	0.25172	12	
	11						0.00000	0.00056	0.01714	0.12752	0.41190	11	
	10					0.00000	0.00001	0.00259	0.04796	0.24466	0.58810	10	
	9				0.00000	0.00001	0.00006	0.00998	0.11338	0.40440	0.74828	9	
	8			0.00000	0.00001	0.00009	0.00042	0.03214	0.22773	0.58411	0.86841	8	
	7			0.00001	0.00011	0.00064	0.00239	0.08869	0.39199	0.74999	0.94234	7	
	6		0.00000	0.00010	0.00087	0.00380	0.01125	0.19570	0.58303	0.87440	0.97931	6	
	5	0.00000	0.00004	0.00096	0.00563	0.01834	0.04317	0.37035	0.76249	0.94905	0.99409	5	
	4	0.00004	0.00060	0.00741	0.02897	0.07062	0.13295	0.58855	0.89291	0.98404	0.99871	4	
	3	0.00100	0.00707	0.04386	0.11497	0.21205	0.32307	0.79302	0.96452	0.99639	0.99980	3	
	2	0.01686	0.05990	0.18966	0.33955	0.48314	0.60825	0.93032	0.99236	0.99948	0.99993	2	
	1	0.18209	0.33239	0.55800	0.70989	0.81131	0.87842	0.93847	0.99920	0.99996	1.00000	1	

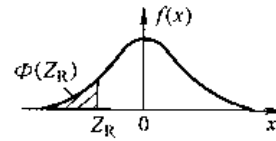
(续)

$$\sum_{r=k}^n \binom{n}{r} p^r (1-p)^{n-r}$$

n	k	p										k	n	
		0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.3	0.4	0.5			
25	25													25
	24										0.00000			24
	23										0.00001			23
	22									0.00000	0.00008			22
	21									0.00001	0.00046			21
	20									0.00005	0.00204			20
	19								0.00000	0.00028	0.00732			19
	18								0.00002	0.00121	0.02164			18
	17								0.00010	0.00433	0.05388			17
	16							0.00000	0.00045	0.01317	0.11476			16
	15							0.00001	0.00178	0.03439	0.21218			15
	14							0.00008	0.00599	0.07780	0.34502			14
	13							0.00037	0.01747	0.15377	0.50000			13
	12						0.00000	0.00154	0.04425	0.26772	0.65498			12
	11						0.00000	0.00001	0.00556	0.09780	0.41423	0.78782		11
10				0.00000	0.00001	0.00008	0.01733	0.18944	0.57538	0.88524			10	
9				0.00001	0.00008	0.00046	0.04677	0.32307	0.72647	0.94612			9	
8			0.00000	0.00007	0.00052	0.00226	0.10912	0.48815	0.84645	0.97836			8	
7		0.00000	0.00004	0.00051	0.00277	0.00948	0.21996	0.65935	0.92643	0.99248			7	
6		0.00001	0.00038	0.00306	0.01229	0.03340	0.38331	0.80651	0.97064	0.99796			6	
5	0.00000	0.00012	0.00278	0.01505	0.04514	0.09799	0.57933	0.90953	0.99053	0.99954			5	
4	0.00011	0.00145	0.01652	0.05976	0.13509	0.23641	0.76601	0.96676	0.99763	0.99992			4	
3	0.00195	0.01324	0.07648	0.18711	0.32317	0.46291	0.90177	0.99104	0.99957	0.99999			3	
2	0.02576	0.08365	0.26419	0.44734	0.60528	0.72879	0.97261	0.99843	0.99995	1.00000			2	
1	0.22218	0.39654	0.63960	0.78709	0.87564	0.92821	0.99622	0.99987	1.00000	1.00000			1	
30	30													30
	29													29
	28													28
	27										0.00000			27
	26										0.00003			26
	25									0.00000	0.00016			25
	24									0.00001	0.00072			24
	23									0.00005	0.00261			23
	22							0.00000	0.00022	0.00806				22
	21							0.00001	0.00086	0.02139				21
	20							0.00004	0.00285	0.04937				20
	19							0.00016	0.00830	0.10024				19
	18							0.00000	0.00063	0.02124	0.18080			18
	17							0.00001	0.00212	0.04811	0.29233			17
	16							0.00005	0.00637	0.09706	0.42777			16
15							0.00023	0.01694	0.17537	0.57223			15	
14							0.00090	0.04005	0.28550	0.70767			14	
13						0.00000	0.00311	0.08147	0.42153	0.81920			13	
12					0.00000	0.00002	0.00949	0.15932	0.36891	0.89976			12	
11				0.00000	0.00001	0.00009	0.02562	0.26963	0.70853	0.95063			11	
10				0.00001	0.00007	0.00045	0.06109	0.41119	0.82371	0.97801			10	
9			0.00000	0.00005	0.00041	0.00202	0.12365	0.56348	0.90599	0.99194			9	
8			0.00002	0.00030	0.00197	0.00778	0.23921	0.71862	0.95648	0.99739			8	
7		0.00000	0.00015	0.00167	0.00825	0.02583	0.39303	0.84048	0.98282	0.99928			7	
6	0.00000	0.00003	0.00106	0.00795	0.02929	0.07319	0.57249	0.92341	0.99434	0.99984			6	
5	0.00001	0.00030	0.00632	0.03154	0.08736	0.17549	0.54477	0.96985	0.99849	0.99997			5	
4	0.00022	0.00289	0.03059	0.10262	0.21579	0.35256	0.87729	0.99068	0.99969	1.00000			4	
3	0.00332	0.02172	0.11690	0.26760	0.43460	0.58865	0.95582	0.99780	0.99995	1.00000			3	
2	0.03815	0.12055	0.33882	0.54453	0.70421	0.81630	0.93948	0.99969	1.00000	1.00000			2	
1	0.20030	0.45452	0.70614	0.84374	0.91803	0.95761	0.99378	1.00000	1.00000	1.00000			1	

表 45.2-4 标准正态分布表

$$R = \Phi(Z_R) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_R} e^{-\frac{x^2}{2}} dx \quad (Z_R \leq 0)$$



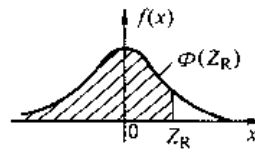
a)

Z <sub>R</sub>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.09853
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.3	0.01072	0.01044	0.01017	0.009903	0.009642	0.009387	0.009137	0.008894	0.008656	0.008424
-2.4	0.008198	0.007976	0.007760	0.007549	0.007344	0.007143	0.006947	0.006756	0.006569	0.006387
-2.5	0.006210	0.006037	0.005868	0.005703	0.005543	0.005386	0.005234	0.005085	0.004940	0.004799
-2.6	0.004661	0.004527	0.004396	0.004269	0.004145	0.004025	0.003907	0.003793	0.003681	0.003573
-2.7	0.003467	0.003364	0.003264	0.003167	0.003072	0.002980	0.002890	0.002803	0.002718	0.002635
-2.8	0.002555	0.002477	0.002401	0.002327	0.002256	0.002186	0.002118	0.002052	0.001988	0.001926
-2.9	0.001866	0.001807	0.001750	0.001695	0.001641	0.001589	0.001538	0.001489	0.001441	0.001395
-3.0	0.001350	0.001306	0.001264	0.001223	0.001183	0.001144	0.001107	0.001070	0.001035	0.001001
-3.1	0.0009676	0.0009354	0.0009043	0.0008740	0.0008447	0.0008164	0.0007888	0.0007622	0.0007364	0.0007114
-3.2	0.0006871	0.0006637	0.0006410	0.0006190	0.0005976	0.0005770	0.0005571	0.0005377	0.0005190	0.0005009
-3.3	0.0004834	0.0004665	0.0004501	0.0004342	0.0004189	0.0004041	0.0003897	0.0003758	0.0003624	0.0003495
-3.4	0.0003369	0.0003248	0.0003131	0.0003018	0.0002909	0.0002803	0.0002701	0.0002602	0.0002507	0.0002415
-3.5	0.0002326	0.0002241	0.0002158	0.0002078	0.0002001	0.0001926	0.0001854	0.0001785	0.0001718	0.0001653
-3.6	0.0001591	0.0001531	0.0001473	0.0001417	0.0001363	0.0001311	0.0001261	0.0001213	0.0001166	0.0001121
-3.7	0.0001078	0.0001036	0.0000991	0.00009574	0.00009201	0.00008842	0.00008496	0.00008162	0.00007841	0.00007532
-3.8	0.00007235	0.00006948	0.00006673	0.00006407	0.00006152	0.00005906	0.00005669	0.00005442	0.00005223	0.00005012
-3.9	0.00004810	0.00004615	0.00004427	0.00004247	0.00004074	0.00003908	0.00003747	0.00003594	0.00003446	0.00003304
-4.0	0.00003167	0.00003036	0.00002910	0.00002789	0.00002673	0.00002561	0.00002454	0.00002351	0.00002252	0.00002157
-4.1	0.00002066	0.00001978	0.00001894	0.00001814	0.00001737	0.00001662	0.00001591	0.00001523	0.00001458	0.00001395
-4.2	0.00001335	0.00001277	0.00001222	0.00001168	0.00001118	0.00001069	0.00001022	0.000009774	0.000009345	0.000008934
-4.3	0.000008540	0.000008163	0.000007801	0.000007455	0.000007124	0.000006807	0.000006503	0.000006212	0.000005934	0.000005668
-4.4	0.000005413	0.000005169	0.000004935	0.000004712	0.000004498	0.000004294	0.000004098	0.000003911	0.000003732	0.000003561
-4.5	0.000003398	0.000003241	0.000003092	0.000002949	0.000002813	0.000002682	0.000002558	0.000002439	0.000002325	0.000002216
-4.6	0.000002212	0.000002013	0.000001919	0.000001828	0.000001742	0.000001660	0.000001581	0.000001506	0.000001434	0.000001366
-4.7	0.000001301	0.000001239	0.000001179	0.000001123	0.000001069	0.000001017	0.0000009680	0.0000009211	0.0000008765	0.0000008339
-4.8	0.0000007933	0.0000007547	0.0000007178	0.0000006827	0.0000006492	0.0000006173	0.0000005869	0.0000005580	0.0000005304	0.0000005042
-4.9	0.0000004792	0.0000004554	0.0000004327	0.0000004111	0.0000003996	0.0000003711	0.0000003525	0.0000003348	0.0000003179	0.0000003019



(续)

$$K = \Phi(Z_R) = -\frac{1}{\sqrt{2\pi}} \int_{-x}^{Z_R} e^{-\frac{t^2}{2}} dt \quad (Z_R \geq 0)$$



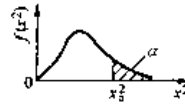
b)

Z <sub>R</sub>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9014
1.3	0.9032	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.5	0.9331	0.9348	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440
1.6	0.9452	0.9463	0.9473	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9544
1.7	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9608	0.9616	0.9624	0.9632
1.8	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9712	0.9719	0.9725	0.9732	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.0	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.1	0.9821	0.9825	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9853	0.9857
2.2	0.9861	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9884	0.9887	0.9890
2.3	0.9892	0.9895	0.9898	0.9901	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.4	0.9918	0.9920	0.9922	0.9924	0.9925	0.9926	0.9927	0.9928	0.9929	0.9930
2.5	0.9931	0.9932	0.9933	0.9934	0.9935	0.9936	0.9937	0.9938	0.9939	0.9940
2.6	0.9941	0.9942	0.9943	0.9944	0.9945	0.9946	0.9947	0.9948	0.9949	0.9950
2.7	0.9951	0.9952	0.9953	0.9954	0.9955	0.9956	0.9957	0.9958	0.9959	0.9960
2.8	0.9961	0.9962	0.9963	0.9964	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970
2.9	0.9971	0.9972	0.9973	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980
3.0	0.9981	0.9982	0.9983	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989	0.9990
3.1	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.2	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.3	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.4	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.5	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.6	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.7	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.8	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
3.9	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.0	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.1	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.2	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.3	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.4	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.5	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.6	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.7	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.8	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000
4.9	0.9991	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9998	0.9999	1.0000

注 1. 0.9<sup>30</sup> = 0.9990, 其余类似。  
 2. 0.0<sup>31</sup> = 0.0001, 其余类似。

表 45.2-5  $\chi^2$  分布表

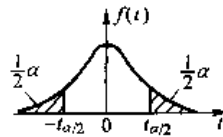
$P(\chi^2 \geq \chi_{\alpha}^2) = \alpha$



$\nu \backslash \alpha$	0.995	0.99	0.98	0.80	0.70	0.30	0.20	0.10	0.05	0.01
1	0.00016	0.00393	0.0158	0.0642	0.148	1.074	1.642	2.706	3.841	6.635
2	0.0201	0.1026	0.211	0.446	0.713	2.408	3.219	4.605	5.991	9.210
3	0.115	0.352	0.584	1.005	0.424	3.665	4.642	6.251	7.815	11.341
4	0.297	0.711	1.064	1.649	2.195	4.878	5.989	7.779	9.488	13.277
5	0.554	1.145	1.610	2.343	3.000	6.064	7.289	9.236	11.070	15.086
6	0.872	1.635	2.204	3.070	3.828	7.231	8.558	10.645	12.592	16.812
7	1.239	2.167	2.833	3.822	4.671	8.383	9.803	12.017	14.067	18.475
8	1.646	2.733	3.490	4.594	5.527	9.524	11.030	13.362	15.507	20.090
9	2.088	3.325	4.168	5.380	6.393	10.656	12.242	14.684	16.919	21.666
10	2.558	3.940	4.865	6.179	7.267	11.781	13.442	15.987	18.307	23.209
11	3.053	4.575	5.578	6.980	8.148	12.899	14.631	17.275	19.675	24.725
12	3.571	5.226	6.304	7.807	9.034	14.011	15.812	18.549	21.026	26.217
13	4.107	5.892	7.042	8.634	9.926	15.119	16.985	19.812	22.362	27.688
14	4.660	6.571	7.790	9.467	10.821	16.222	18.151	21.064	23.685	29.141
15	5.229	7.261	8.547	10.307	11.721	17.322	19.311	22.307	24.996	30.578
16	5.812	7.962	9.312	11.152	12.624	18.418	20.465	23.542	26.296	32.000
17	6.408	8.672	10.085	12.002	13.531	19.511	21.615	24.769	27.587	33.409
18	7.015	9.390	10.865	12.857	14.440	20.601	22.760	25.989	28.869	34.805
19	7.633	10.117	11.651	13.716	15.352	21.689	23.900	27.204	30.144	36.191
20	8.260	10.851	12.443	14.578	16.266	22.775	25.038	28.412	31.410	37.566
21	8.897	11.591	13.240	15.445	17.182	23.858	26.171	29.615	32.671	38.932
22	9.542	12.338	14.042	16.314	18.101	24.939	27.301	30.813	33.924	40.289
23	10.196	13.091	14.848	17.187	19.021	26.018	28.429	32.007	35.172	41.638
24	10.856	13.848	15.659	18.062	19.943	27.096	29.558	33.196	36.415	42.980
25	11.524	14.611	16.473	18.940	20.867	28.172	30.675	34.382	37.652	44.314
26	12.198	15.379	17.292	19.820	21.792	29.246	31.765	35.566	38.885	45.642
27	12.879	16.151	18.114	20.703	22.719	30.319	32.912	36.741	40.113	46.933
28	13.565	16.928	18.939	21.588	23.647	31.391	34.027	37.916	41.337	48.278
29	14.257	17.708	19.768	22.475	24.577	32.461	35.139	39.087	42.557	49.588
30	14.954	18.493	20.599	23.364	25.508	33.530	36.250	40.256	43.773	50.892
40	22.164	26.509	29.051	32.352	34.876	44.163	37.263	51.805	55.758	63.691
60	37.485	43.188	46.459	50.647	53.815	65.225	38.969	74.397	79.082	88.379
80	53.540	60.391	64.278	69.213	72.920	86.122	90.403	96.578	101.879	112.329
100	70.065	77.929	82.358	89.950	92.137	106.908	111.667	118.498	123.342	135.807
200	156.432	168.279	174.835	183.006	189.052	209.997	216.618	226.021	233.994	249.445

表 45.2-6  $t$  分布的双侧分位数 ( $t_{\frac{\alpha}{2}}$ ) 表

$P(|t| > t_{\frac{\alpha}{2}}) = \alpha$



( ) 中  $\alpha$  值是单侧分位数 ( $t_{\alpha}$ )

$\nu \backslash \alpha$	0.9 (0.45)	0.8 (0.4)	0.7 (0.35)	0.6 (0.3)	0.5 (0.25)	0.4 (0.2)	0.3 (0.15)	0.2 (0.1)	0.1 (0.05)	0.05 (0.12)	0.02 (0.01)	0.01 (0.005)	0.001 (0.0005)	$\alpha \nu$
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619	1
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598	2
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.633	2.353	3.182	4.541	5.841	12.924	3
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610	4
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859	5
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959	6

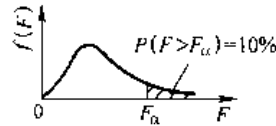
(续)

$\nu$	0.9 (0.45)	0.8 (0.4)	0.7 (0.35)	0.6 (0.3)	0.5 (0.25)	0.4 (0.2)	0.3 (0.15)	0.2 (0.1)	0.1 (0.05)	0.05 (0.12)	0.02 (0.01)	0.01 (0.005)	0.001 (0.0005)	$\alpha$
7	0.130	0.268	0.402	0.540	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405	7
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041	8
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781	9
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587	10
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437	11
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.631	3.055	4.318	12
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221	13
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140	14
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.161	2.602	2.947	4.073	15
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015	16
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965	17
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922	18
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.833	19
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850	20
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.030	2.518	2.831	3.819	21
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792	22
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767	23
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.313	1.711	2.064	2.492	2.797	3.745	24
25	0.127	0.256	0.390	0.531	0.681	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725	25
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707	26
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690	27
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674	28
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659	29
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	0.310	1.697	2.042	2.457	2.750	3.646	30
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551	40
60	0.126	0.254	0.387	0.527	0.670	0.848	1.046	1.293	1.671	2.000	2.390	2.660	3.460	60
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373	120
$\infty$	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291	$\infty$

表 45.2-7 F 分布表

$P(F > F_{\alpha}) = \alpha$

$(\alpha = 0.10)$



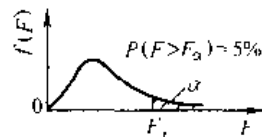
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	61.22	61.74	62.26	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.44	9.46	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.20	5.18	5.17	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.82	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.24	3.21	3.17	3.14	3.12	3.13
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.84	2.80	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.59	2.56	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.42	2.38	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.30	2.25	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.20	2.16	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.17	2.12	2.08	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.10	2.06	2.01	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.05	2.01	1.96	1.90	1.88	1.85
14	3.10	2.78	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.01	1.96	1.91	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.92	1.87	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.94	1.89	1.84	1.78	1.75	1.72
17	3.03	2.65	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.91	1.86	1.81	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.89	1.84	1.78	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.86	1.81	1.76	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.84	1.79	1.74	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.83	1.78	1.72	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.81	1.76	1.70	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.80	1.74	1.69	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.78	1.73	1.67	1.61	1.57	1.53

(续)

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.77	1.72	1.66	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.76	1.71	1.65	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.75	1.70	1.64	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.74	1.69	1.63	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.73	1.68	1.62	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.72	1.67	1.61	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.66	1.61	1.54	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.60	1.54	1.48	1.40	1.35	1.29
120	2.75	2.65	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.55	1.48	1.41	1.32	1.26	1.19
$\infty$	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.49	1.42	1.34	1.24	1.17	1.00

$P(F > F_\alpha) = \alpha$

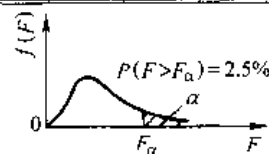
( $\alpha = 0.05$ )



$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
1	161	200	216	225	230	234	237	239	241	242	246	248	250	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.65	3.51	3.44	3.38	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.22	3.15	3.08	3.00	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.84	2.77	2.70	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.57	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.47	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.38	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.31	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.19	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.15	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.11	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.07	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.18	2.10	2.01	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	1.98	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05	1.96	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.94	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.92	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.90	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.06	1.97	1.88	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.87	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.03	1.94	1.85	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.55	1.43	1.35	1.25
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.32	1.22	1.00

$P(F > F_\alpha) = \alpha$

( $\alpha = 0.025$ )



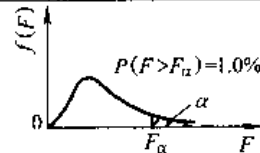
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
1	648	800	864	900	922	937	948	957	963	969	985	993	1001	1010	1014	1018
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5

(续)

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
3	17.4	16.0	15.4	15.1	14.9	14.7	14.4	14.5	14.5	14.4	14.3	14.2	14.1	14.0	13.9	13.9
4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.36	8.31	8.26
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.33	3.23	3.12	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.39	3.18	3.07	2.96	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.05	2.95	2.84	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.95	2.84	2.73	2.61	2.55	2.49
15	6.20	4.76	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.79	2.68	2.57	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.96	2.72	2.62	2.50	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.67	2.56	2.44	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.62	2.51	2.39	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.53	2.42	2.31	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.50	2.39	2.27	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.47	2.36	2.24	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.44	2.33	2.21	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.41	2.30	2.18	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.39	2.28	2.16	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.36	2.25	2.13	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.34	2.23	2.11	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.32	2.21	2.09	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.94	1.82	1.69	1.53	1.43	1.31
$\infty$	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.39	1.27	1.00

$P(F > F_\alpha = \alpha)$

( $\alpha = 0.01$ )



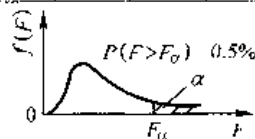
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
1	4052	5000	5403	5625	5764	5859	5928	5982	6002	6056	6157	6209	6261	6313	6339	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	26.9	26.7	26.5	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.2	14.0	13.8	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.72	9.55	9.38	9.20	9.11	9.08
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.23	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	5.99	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.20	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.65	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.44	4.25	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	3.94	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.70	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.51	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.35	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.21	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.26	3.10	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.31	3.16	3.00	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	3.08	2.92	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.15	3.00	2.84	2.67	2.58	2.49

(续)

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.78	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.03	2.88	2.72	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.83	2.67	2.50	2.40	2.31
23	7.88	5.66	4.76	4.21	3.94	3.71	3.54	3.41	3.30	3.21	2.93	2.78	2.62	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.89	2.74	2.58	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.85	2.70	2.54	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.82	2.66	2.50	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.78	2.63	2.47	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.75	2.60	2.44	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.73	2.57	2.41	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.39	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.20	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.20	2.03	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.19	2.03	1.86	1.66	1.53	1.38
$\infty$	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.04	1.88	1.70	1.47	1.32	1.00

$P(F > F_\alpha) = \alpha$

( $\alpha = 0.005$ )



$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	60	120	$\infty$
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24630	24836	25044	25253	25359	25465
2	198	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
3	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.1	42.8	42.5	42.2	42.0	41.8
4	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.4	20.2	19.9	19.6	19.5	19.3
5	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.1	12.9	12.7	12.4	12.3	12.1
6	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	9.81	9.59	9.36	9.12	9.00	8.88
7	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	7.97	7.75	7.53	7.31	7.19	7.08
8	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	6.81	6.61	6.40	6.18	6.06	5.95
9	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.03	5.83	5.62	5.41	5.30	5.19
10	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.47	5.27	5.07	4.86	4.75	4.64
11	12.2	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.05	4.86	4.65	4.44	4.34	4.23
12	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.73	4.53	4.34	4.12	4.01	3.90
13	11.4	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.46	4.27	4.07	3.87	3.76	3.65
14	11.1	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.25	4.06	3.86	3.66	3.55	3.44
15	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.07	3.88	3.69	3.48	3.37	3.26
16	10.6	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	3.92	3.73	3.54	3.33	3.22	3.11
17	10.4	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.79	3.61	3.41	3.21	3.10	2.98
18	10.2	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.68	3.50	3.30	3.10	2.99	2.87
19	10.1	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.59	3.40	3.21	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.50	3.32	3.12	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.43	3.24	3.05	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.36	3.18	2.98	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.30	3.12	2.92	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.25	3.06	2.87	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.20	3.01	2.82	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.15	2.97	2.77	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.11	2.93	2.73	2.52	2.41	2.29
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.07	2.89	2.69	2.48	2.37	2.25
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.04	2.86	2.66	2.45	2.33	2.21
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.01	2.82	2.63	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.78	2.60	2.40	2.18	2.06	1.93
60	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.12	3.01	2.90	2.57	2.39	2.19	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.37	2.19	1.98	1.75	1.61	1.43
$\infty$	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.19	2.00	1.79	1.53	1.36	1.00

## 2 假设检验

数理统计是进行数据处理和分析的基础。可靠性设计主要涉及的两类问题是分布类型的假设检验和母体参数估计。

数理统计常用的名词术语有：

母体——所研究对象的全体称为母体，也叫总体；

个体——组成母体的每个基本单元称为个体；

子样——为了研究母体而以适当的方法在母体中抽取一部分个体，称为子样或样本；

子样容量——子样所包含的个体数目,称为子样容量,记为  $n$ ;

随机抽样——不掺入人为的主观因素而具有随机性的抽样,称为随机抽样;

统计量——根据子样数据计算得到的参数称为统计量,例如子样均值、子样标准差等;

自由度——统计量中所包含无约束自变量的个数,称为该统计量的自由度。例如,统计量  $S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  中,有一个约束  $x_1 + x_2 + \dots + x_n - n\bar{x} = 0$ ,故  $S_x^2$  的自由度为  $n-1$ 。

### 2.1 分布类型的假设检验

随机变量的分布类型一般是需要首先确定的。对分布不明的情况,则应做大子样的试验判定分布类型。对已有经验参考的,则可做较小子样的试验,假设其分布类型后,再进行相应的拟合性检验。

检验方法大致可分两类,一类是数值分析法,另一类是图分析法。图分析法是使用各种坐标纸进行分析,这种方法简便易行、直观易懂。缺点是难以完全避免主观因素的影响,其结果往往因人而异。数值分析法精确性高且具有客观性,但计算过程较为复杂。随着计算机科学技术的发展,目前已有多种大型可靠性分析软件包,为数值分析法提供了良好的工具。

下面先介绍数值分析法中最常用的  $\chi^2$  检验法和 K-S 检验法,再对几种最常用的分布类型给出些专用的检验法。

### 2.2 $\chi^2$ 检验法

$\chi^2$  检验法使用范围很广,可用于离散随机变量和连续随机变量,母体参数已知或未知,全子样或截尾等。但  $\chi^2$  检验法一般只用于大子样,子样值个数至少大于 50,最好在 100 以上。对于截尾子样, $\chi^2$  检验法容易犯第 II 类错误,即接受了不正确的原假设。

$\chi^2$  检验的步骤是:①计算理论频数与实际频数间的差异;②将检验统计量  $\chi^2$  的观测值与临界值  $\chi^2_\alpha(k-m-1)$  比较,满足下式条件,则接受原假设,否则,拒绝原假设。

$$\chi^2 = \sum_{i=1}^k \frac{(v_i - np_i)^2}{np_i} \leq \chi^2_\alpha(k-m-1) \tag{45.2-26}$$

式中  $n$ ——子样容量;

$k$ ——分组数,按子样容量取  $k=7\sim 14$ ;

$v_i$ ——第  $i$  组实际频数,一般宜取  $v_i \geq 5$ ;

$p_i$ ——第  $i$  组理论频率(概率);

$m$ ——未知参数(需估计的参数)数目;

$\alpha$ ——显著性水平;

$\chi^2_\alpha(\cdot)$ —— $\chi^2$  临界值,由  $\chi^2$  分布表(表 45.2-5)查得。

例 观察 200 个某种产品的失效时间,每隔 100h 检查一次,记下失效产品个数,直到全部失效为止。记录如表 45.2-8 所示,试检验该产品寿命是否服从指数分布。

解 假设寿命服从指数分布,但参数  $\lambda$  未知,需先估计。取各组的中值为该组代表值  $t_i$ ,计算子样均值  $\bar{t}$  故

$$\begin{aligned} \bar{t} &= \frac{1}{n} \sum_{i=1}^k t_i v_i = \\ &= \frac{1}{200} (50 \times 36 + 150 \times 42 + 250 \times 37 \\ &\quad + 350 \times 23 + 450 \times 20 + 550 \times 25 + \\ &\quad + 650 \times 10 + 750 \times 4 + 850 \times 3) h = 301 h \\ \hat{\lambda} &= \frac{1}{\bar{t}} = \frac{1}{301} 1/h \end{aligned}$$

假设  $H_0: F(t) = 1 - e^{-\frac{t}{301}}$

应用  $\chi^2$  检验法,首先按规定分组,由于每组中实际频数不宜少于 5,故将前 7 段时间各作为一组,最后两段时间合为一组,故总计组数  $k=8$ ,正好在 7~14 范围内。计算结果见表 45.2-9。

表 45.2-8 数据记录

时间 /h	0~100	~200	~300	~400	~500	~600	~700	~800	~900
失效数 $r_i$	36	42	37	23	20	25	10	4	3

表 45.2-9 计算结果

组号 $i$	$v_i = r_i$	$p_i = \left(1 - e^{-\frac{t_i}{301}}\right) - \left(1 - e^{-\frac{t_{i-1}}{301}}\right)$	$np_i = 200p_i$	$v_i - np_i$	$(v_i - np_i)^2$	$\frac{(v_i - np_i)^2}{np_i}$
1	36	0.2827	56.54	-20.54	421.8916	7.4618
2	42	0.2028	40.56	1.44	2.0736	0.0511
3	37	0.1455	29.10	7.90	62.4100	2.1447
4	23	0.1043	20.86	2.14	4.5796	0.2195
5	20	0.0749	14.98	5.02	25.2004	1.6823
6	25	0.0537	10.74	14.26	203.3476	18.9337
7	10	0.0385	7.70	2.30	5.2900	0.6870
8	7	0.0474	9.48	-2.48	6.1504	0.6488
$\Sigma$						31.8352

根据计算结果,  $\chi^2 = \sum_{i=1}^k \frac{(v_i - np_i)^2}{np_i} = 31.8352$ 。

取显著性水平  $\alpha=0.10$ , 由  $\chi^2$  分布表查得

$$\chi^2_{\alpha}(k - m - 1) = \chi^2_{0.10}(8 - 1 - 1) = 10.64$$

由于  $\chi^2 > \chi^2_{0.10}(6)$ , 故不能接受原假设, 即不能认为该产品寿命服从指数分布。

### 2.3 K-S 检验法

K-S 检验法 ( $d$  检验) 比  $\chi^2$  检验法精细, 而且还适用于小子样的情况。但是, K-S 检验法要求所检验分布中不含未知参数。当指定分布中含有未知参数时, 对某些分布应该用专门的临界值表。

K-S 检验的步骤是: ①将  $n$  个试验数据按由小到大的次序排列; ②根据假设的分布, 计算每个数据对应的  $F_0(x_i)$ ; ③将其与经验分布函数  $F_n(x_i)$  相比较, 其差的最大绝对值即检验统计量  $D_n$  的观察值, 将  $D_n$  与临界值  $D_{n,\alpha}$  比较。满足下式条件, 则接受原假设, 否则拒绝原假设。

$$D_n = \max_{-0.5 \leq x \leq 0.5} |F_n(x_i) - F_0(x_i)| = \max\{d_i\} \leq D_{n,\alpha} \quad (45.2-27)$$

式中  $F_0(x_i)$  —— 原假设分布函数;

$F_n(x_i)$  —— 子样容量为  $n$  的经验分布函数;

$$F_n(x_i) = \begin{cases} 0 & x < x_1 \\ \frac{i}{n} & x_1 \leq x \leq x_{i+1} \\ 1 & x \geq x_n \end{cases} \quad x_1 \leq x_2 \leq \dots \leq x_n \quad (45.2-28)$$

$$d_i = \max \left\{ F_n(x_i) - \frac{i-1}{n}, \frac{i}{n} - F_0(x_i) \right\} \quad (45.2-29)$$

$D_{n,\alpha}$  —— 临界值, 见表 45.2-10。

表 45.2-10 K-S 检验临界值 ( $D_{n,\alpha}$ ) 表

$n$	$\alpha$	0.20	0.10	0.05	0.02	0.01
1		0.90000	0.95000	0.97500	0.99000	0.99500
2		0.68377	0.77639	0.84189	0.90000	0.92929
3		0.56481	0.63604	0.70760	0.78456	0.82900
4		0.49265	0.56522	0.62394	0.68887	0.73424
5		0.44698	0.50945	0.56328	0.62718	0.66853
6		0.41037	0.46799	0.51926	0.57741	0.61661
7		0.38148	0.43607	0.48342	0.53844	0.57581
8		0.35831	0.40962	0.45427	0.50654	0.54179
9		0.33910	0.38746	0.43001	0.47960	0.51332
10		0.32260	0.36866	0.40925	0.45662	0.48893
11		0.30829	0.35242	0.39122	0.43670	0.46770
12		0.29577	0.33815	0.37543	0.41918	0.44905
13		0.28470	0.32549	0.36143	0.40362	0.43247
14		0.27481	0.31417	0.34890	0.38970	0.41762
15		0.26588	0.30397	0.33760	0.37713	0.40420
16		0.25778	0.29472	0.32733	0.36571	0.39201
17		0.25039	0.28627	0.31796	0.35528	0.38086

(续)

$n$	$\alpha$	0.20	0.10	0.05	0.02	0.01
18		0.24360	0.27851	0.30936	0.34569	0.37062
19		0.23735	0.27136	0.30143	0.33685	0.36117
20		0.23156	0.26473	0.29408	0.32866	0.35241
21		0.22617	0.25858	0.28724	0.32104	0.34427
22		0.22115	0.25283	0.28087	0.31394	0.33666
23		0.21645	0.24740	0.27490	0.30728	0.32954
24		0.21205	0.24242	0.26931	0.30104	0.32286
25		0.20790	0.23768	0.26404	0.29516	0.31657
26		0.20399	0.23320	0.25907	0.28962	0.31064
27		0.20030	0.22898	0.25438	0.28438	0.30502
28		0.19680	0.22497	0.24993	0.27942	0.29971
29		0.19348	0.22117	0.24571	0.27471	0.29466
30		0.19032	0.21756	0.24170	0.27023	0.28987
31		0.18732	0.21412	0.23788	0.26596	0.28530
32		0.18445	0.21085	0.23424	0.26189	0.28094
33		0.18171	0.20771	0.23076	0.25801	0.27677
34		0.17909	0.20472	0.22743	0.25429	0.27279
35		0.17659	0.20185	0.22425	0.25073	0.26897
36		0.17418	0.19910	0.22119	0.24732	0.26532
37		0.17188	0.19646	0.21826	0.24404	0.26180
38		0.16966	0.19392	0.21544	0.24089	0.25843
39		0.16753	0.19148	0.21273	0.23786	0.25518
40		0.16547	0.18913	0.21012	0.23494	0.25205
41		0.16349	0.18687	0.20760	0.23212	0.24904
42		0.16158	0.18468	0.20517	0.22941	0.24613
43		0.15974	0.18257	0.20283	0.22679	0.24332
44		0.15796	0.18053	0.20056	0.22426	0.24060
45		0.15623	0.17856	0.19837	0.22181	0.23798
46		0.15457	0.17665	0.19625	0.21944	0.23544
47		0.15295	0.17481	0.19420	0.21715	0.23298
48		0.15139	0.17302	0.19221	0.21493	0.23059
49		0.14987	0.17128	0.19028	0.21277	0.22828
50		0.14840	0.16959	0.18841	0.21068	0.22604
55		0.14164	0.16186	0.17981	0.20107	0.21574
60		0.13573	0.15511	0.17231	0.19267	0.20673
65		0.13052	0.14913	0.16567	0.18525	0.19877
70		0.12586	0.14381	0.15975	0.17863	0.19167
75		0.12167	0.13901	0.15442	0.17268	0.18528
80		0.11787	0.13467	0.14960	0.16728	0.17949
85		0.11442	0.13072	0.14520	0.16236	0.17421
90		0.11125	0.12709	0.14117	0.15786	0.16938
95		0.10833	0.12375	0.13746	0.15371	0.16493
100		0.10563	0.12067	0.13403	0.14987	0.16081

例 某合金的 9 个试样测得的强度极限为 453MPa, 436MPa, 429MPa, 419MPa, 405MPa, 416MPa, 432MPa, 423MPa, 440MPa。检验该合金的强度极限是否服从均值  $\mu=428$ MPa, 标准差  $\sigma=15$ MPa 的正态分布。

解 将该合金的强度极限  $X$  的数据按由小到大的次序排列。假设  $X$  服从正态分布, 分布函数  $F(x) = \int_{-\infty}^x \frac{1}{15\sqrt{2\pi}} e^{-\frac{(x-428)^2}{2 \times 15^2}} dx = \Phi\left(\frac{x-428}{15}\right)$  式中的  $\Phi(\cdot)$  查表 45.2-4, 计算结果见表 45.2-11。根据式 (45.2-27)  $D_n$  的观察值为

$$D_n = \max\{d_i\} = 0.1009$$

取显著性水平  $\alpha=0.10$ , 由表 45.2-10 查得  $D_{n,\alpha} = 0.38746$ , 由于  $D_n < D_{n,\alpha}$ , 故接受原假设, 即认为该合金的强度极限服从  $\mu=428$ MPa,  $\sigma=15$ MPa 的正态分



布。

表 45.2-11 算例所列计算表

序号 $i$	$x_i$	$F_0(x_i) = \Phi\left(\frac{x_i - 428}{15}\right)$	$\frac{i-1}{n}$	$\frac{i}{n}$	$d_i$
1	405	0.06301	0	0.111	0.06301
2	416	0.2119	0.111	0.222	0.1009
3	419	0.2743	0.222	0.333	0.0587
4	423	0.3707	0.333	0.444	0.0733
5	429	0.5279	0.444	0.556	0.0839
6	432	0.6064	0.556	0.667	0.0606
7	436	0.7019	0.667	0.778	0.0761
8	440	0.7881	0.778	0.889	0.1009
9	453	0.95254	0.889	1.000	0.06354

### 3 分布参数的估计

分布参数的估计就是根据子样的试验数据来估计已知的母体分布函数中的未知参数。由于母体的分布函数为已知,因此分布参数的估计包括点估计和区间估计。必要时对有关参数进行相关性回归分析。

#### 3.1 点估计

对母体参数的点估计,是用统计量的单一数值去估计母体一个未知参数的数值。如果  $x$  是具有概率分布  $f(x)$  的随机变量,子样容量为  $n$ ,子样值为  $x_1, x_2, \dots, x_n$ ,则与其未知参数相应的统计量  $\hat{\theta} = h(x_1, x_2, \dots, x_n)$ ,称为  $\theta$  的估计值。因为  $\hat{\theta}$  是子样的函数,所以是一个随机变量。在子样确定之后,就能得到一个确定的  $\hat{\theta}$  值,这就是  $\theta$  的点估计。

例如,设随机变量  $x$  服从正态分布,其母体均值  $\mu$  和方差  $\sigma^2$  为未知。子样均值  $\bar{x}$  是未知的母体均值  $\mu$  的点估计,即  $\hat{\mu} = \bar{x}$ 。在子样选好之后,由子样数据算出的数值  $\bar{x}$  是  $\mu$  的点估计。同理,子样方差  $s^2$  是母体方差  $\sigma^2$  的点估计,即  $\hat{\sigma}^2 = s^2$ 。

#### 3.2 区间估计

通常,所作的参数点估计与母体参数有一定的差异。区间估计是要估计母体真值所在的区间,而且区间范围越小越好。

如果取子样的两个统计量  $\hat{\theta}_L(x_1, x_2, \dots, x_n)$ , 和  $\hat{\theta}_U(x_1, x_2, \dots, x_n)$ ,使区间  $[\hat{\theta}_L, \hat{\theta}_U]$  以某一规定的概率  $(1-\gamma)$  包含未知参数  $\theta$ ,这就是区间估计,表示为

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \gamma \quad (45.2-30)$$

式中  $\gamma$  — 风险。

也可以这样表述:用  $\hat{\theta}$  作为未知参数  $\theta$  的估计值,则  $\hat{\theta}$  与真值  $\theta$  的误差不超过某一正数  $\epsilon$  的概率为  $1-\gamma$ ,记为

$$P\{|\hat{\theta} - \theta| < \epsilon\} = 1 - \gamma$$

或 
$$P\{\hat{\theta} - \epsilon < \theta < \hat{\theta} + \epsilon\} = 1 - \gamma \quad (45.2-31)$$

也就是  $\theta$  位于区间  $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$  内的概率。通常把这个概率  $(1-\gamma)$  称为置信度,而把区间  $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$  或  $[\hat{\theta}_L, \hat{\theta}_U]$  称为置信区间,置信区间表示估计结果的精确性,而置信度表示估计结果的可靠性。

下面介绍的是两种常见的区间估计问题。

1) 方差  $\sigma^2$  已知,对母体均值  $\mu$  进行区间估计

设已知随机变量  $x$  的方差  $\sigma^2$ ,子样容量  $n$ ,子样值  $x_1, x_2, \dots, x_n$ ,则对于母体均值  $\mu$  的置信区间估计可以由子样均值  $\bar{x}$  的抽样分布得到。

如果随机变量  $x$  为正态或近似正态分布,则子样均值  $\bar{x}$  的抽样分布为正态分布。因此,统计量  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  的分布为一标准正态分布,如图 45.2-1 所示。

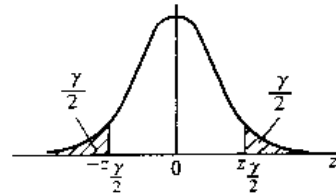


图 45.2-1 统计量  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  的分布

由图 45.2-1 看出,统计量  $z$  介于  $-z_{\gamma/2}$  和  $+z_{\gamma/2}$  之间的概率为

$$P(-z_{\gamma/2} \leq z \leq z_{\gamma/2}) = 1 - \gamma$$

$$P\left[-z_{\gamma/2} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\gamma/2}\right] = 1 - \gamma$$

由此可得

$$P\left(\bar{x} - \frac{z_{\gamma/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{z_{\gamma/2}\sigma}{\sqrt{n}}\right) = 1 - \gamma \quad (45.2-32)$$

在置信度为  $1-\gamma$  时,母体均值  $\mu$  的双侧置信区间为

$$\bar{x} - \frac{z_{\gamma/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{z_{\gamma/2}\sigma}{\sqrt{n}} \quad (45.2-33)$$

上式可以表示为

$$\mu_{L2} \leq \mu \leq \mu_{U2}$$

式中  $\mu$  的下标 L 和 U 分别表示下限和上限,2 表示双侧。显然

$$\left. \begin{aligned} \mu_{L2} &= \bar{x} - \frac{z_{\gamma/2}\sigma}{\sqrt{n}} \\ \mu_{U2} &= \bar{x} + \frac{z_{\gamma/2}\sigma}{\sqrt{n}} \end{aligned} \right\} \quad (45.2-34)$$

单侧置信区间上限和下限分别为

$$\left. \begin{aligned} \mu_{(1)} &\leq \bar{x} + \frac{z_{\gamma}\sigma}{\sqrt{n}} \\ \mu_{(1)} &= \bar{x} - \frac{z_{\gamma}\sigma}{\sqrt{n}} \end{aligned} \right\} \quad (45.2-35)$$

由式(45.2-33)可知,置信区间与置信限都与子样容量  $n$  有关。

例 已知一种钢材强度的标准差  $\sigma=20\text{MPa}$ ,今测得 50 根钢材的强度值,得子样均值  $\bar{x}=700\text{MPa}$ ,如有 5% 的风险,试估计钢材强度的上、下限。

解 当置信度为  $1-\gamma=1-0.05=0.95$  时,相应的标准偏量  $z_{\frac{\gamma}{2}}=1.96$ ,  $\sigma=20\text{MPa}$ ,  $n=50$ ,  $\bar{x}=700\text{MPa}$ ,由式(45.2-34)可得钢材强度的下限值和上限值分别为

$$\bar{x} - \frac{z_{\frac{\gamma}{2}}\sigma}{\sqrt{n}} = \left( 700 - \frac{1.96 \times 20}{\sqrt{50}} \right) \text{MPa} = 694.5\text{MPa}$$

$$\bar{x} + \frac{z_{\frac{\gamma}{2}}\sigma}{\sqrt{n}} = \left( 700 + \frac{1.96 \times 20}{\sqrt{50}} \right) \text{MPa} = 705.5\text{MPa}$$

因此,钢材强度的置信区间为  $[694.5, 705.5]\text{MPa}$ 。

2) 方差  $\sigma^2$  未知,对母体均值  $\mu$  进行区间估计

设随机变量  $x$  服从正态分布,其母体均值  $\mu$  和方差  $\sigma^2$  为未知,已知子样容量  $n$ ,子样值  $x_1, x_2, \dots, x_n$  则对于母体均值  $\mu$  的区间估计可由  $t$  分布求得。

$t$  分布如图 45.2-2 所示。

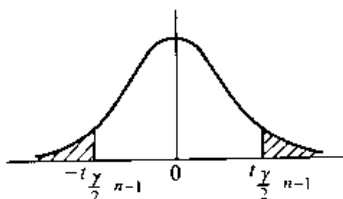


图 45.2-2  $t$  分布

统计量

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad (45.2-36)$$

式中  $s$ ——子样的标准差。

由图 45.2-2 可知

$$P\left(-t_{\frac{\gamma}{2}, n-1} \leq t \leq t_{\frac{\gamma}{2}, n-1}\right) = 1 - \gamma$$

$$\text{或 } P\left[-t_{\frac{\gamma}{2}, n-1} \leq \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \leq t_{\frac{\gamma}{2}, n-1}\right] = 1 - \gamma$$

重新整理后可得

$$P\left[\bar{x} - t_{\frac{\gamma}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\gamma}{2}, n-1} \frac{s}{\sqrt{n}}\right] = 1 - \gamma \quad (45.2-37)$$

式中  $t_{\frac{\gamma}{2}, n-1}$ —— $t$  分布中与  $\frac{\gamma}{2}$  和  $n-1$  相对应的分位数,可由表 45.2-6 查得。

$n-1$ —— $t$  分布的自由度;

$n$ ——子样容量;

$s$ ——子样的标准差。

在置信度为  $1-\gamma$  时,母体均值  $\mu$  的双侧置信区间为

$$\bar{x} - t_{\frac{\gamma}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\gamma}{2}, n-1} \frac{s}{\sqrt{n}} \quad (45.2-38)$$

当为单侧置信区间时,母体均值  $\mu$  的单侧置信区间上限和下限分别为

$$\left. \begin{aligned} \mu_{(1)} &\leq \bar{x} + \frac{t_{\gamma, n-1}s}{\sqrt{n}} \\ \mu_{(1)} &\geq \bar{x} - \frac{t_{\gamma, n-1}s}{\sqrt{n}} \end{aligned} \right\} \quad (45.2-39)$$

例 在制造连杆时,中心孔之间的平均距离  $\bar{x}=21.275\text{mm}$ ,子样容量  $n=6$ ,子样的标准差  $s=0.054\text{mm}$ ,风险为  $\gamma=0.05$ ,求  $\mu$  的置信区间。

解 由式(45.2-39)可得  $\mu$  的置信区间为

$$\begin{aligned} &\left[ \bar{x} - t_{\gamma, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\gamma, n-1} \frac{s}{\sqrt{n}} \right] \\ &= \left[ 21.275 - t_{0.05, 5} \frac{0.054}{\sqrt{6}}, 21.275 + \right. \\ &\quad \left. t_{0.05, 5} \frac{0.054}{\sqrt{6}} \right] \text{mm} \\ &= [21.218, 21.332] \text{mm} \end{aligned}$$

### 3.3 极大似然估计

极大(最大)似然性原理常用于母体参数的点估计量问题。为说明其一般原理,先介绍最简单的特例,设  $X_1$  和  $X_2$  是两个分布相同的正态随机变量  $N(\mu, \sigma^2)$ ,如以  $x_1$  和  $x_2$  分别表示它们的取值,则  $X_1$  和  $X_2$  各自的正态概率密度函数分别为

$$f(x_1 | \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \quad (45.2-40)$$

$$f(x_2 | \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_2 - \mu)^2}{2\sigma^2}} \quad (45.2-41)$$

当  $X_1$  和  $X_2$  互相独立时,由概率乘法定理得二元正态变量的概率函数为

$$\begin{aligned} p(x_1, x_2 | \mu, \sigma^2) &= f(x_1 | \mu, \sigma^2) f(x_2 | \mu, \sigma^2) \\ &= \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^2 \exp \left\{ - \left[ \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2}{2\sigma^2} \right] \right\} \end{aligned}$$

在三维坐标系中,如以  $x_1$  和  $x_2$  表示水平面内的两个坐标,那么,函数  $p(x_1, x_2 | \mu, \sigma^2)$  的几何图形便是一个空间曲面,曲面上任一点的高度对应于平面  $x_1-$

$x_2$  上的一点坐标  $(x_1, x_2)$ 。曲面具体形状取决于  $\mu$  和  $\sigma^2$  之值,因此,可以把这个高度  $p(x_1, x_2 | \mu, \sigma^2)$  看作  $\mu$  和  $\sigma^2$  的函数。今将此函数记作  $L(\mu, \sigma^2 | x_1, x_2)$ , 这就是所谓的“似然函数”。当已经取得  $x_1$  和  $x_2$  的观测值时,似然函数为  $\mu$  和  $\sigma^2$  的二元函数。按以上所述,  $L(\mu, \sigma^2 | x_1, x_2)$  和  $p(x_1, x_2 | \mu, \sigma^2)$  两个函数的形式应该是完全相同的。不过在  $p(x_1, x_2 | \mu, \sigma^2)$  函数中将  $\mu$  和  $\sigma^2$  看作已给定的值,  $x_1$  和  $x_2$  是自变量;而在  $L(\mu, \sigma^2 | x_1, x_2)$  函数中则将  $x_1$  和  $x_2$  看作已知的,  $\mu$  和  $\sigma^2$  是自变量。当把  $\mu$  和  $\sigma^2$  作为自变量时,空间曲面形状将随之发生变化。

极大似然性原理提供一种方法确定  $\mu$  和  $\sigma^2$  之值。那就是使  $L(\mu, \sigma^2 | x_1, x_2)$  取得极大值作为确定  $\mu$  和  $\sigma^2$  的准则。从几何意义上看,要寻求一对数值  $\mu$  和  $\sigma^2$ , 使对应于平面上已给定的一点  $(x_1, x_2)$  的曲面之高度  $p(x_1, x_2 | \mu, \sigma^2)$  达极大值。此种作法的依据,无非是使  $X_1$  在  $x_1$  和  $x_1 + dx_1$  之间并且  $X_2$  在  $x_2$  和  $x_2 + dx_2$  之间同时发生的概率  $p(x_1, x_2 | \mu, \sigma^2) dx_1 dx_2$  最大。今已知似然性函数是两个正态概率函数的乘积

$$L(\mu, \sigma^2 | x_1, x_2) = f(\mu, \sigma^2 | x_1) f(\mu, \sigma^2 | x_2)$$

由于母体参数  $\mu$  和  $\sigma$  相同,所以这两个概率函数可用一条平面曲线来表达(图 45.2-3)。由图可以看到,似然性函数  $L$  实际上等于对应  $x_1$  和  $x_2$  的两个纵坐标之乘积。这样,极大似然性原理意味通过改变曲线的位置( $\mu$ )和形状( $\sigma^2$ ),以使其纵坐标(概率密度)之乘积取得极大值。为了寻求似然性函数  $L$  的极大值,常常求  $L$  自然对数  $\ln L$  的极大值。因为  $L > 0$ ,  $\ln L$  的一阶导数等于零的条件是

$$(\ln L)' = \frac{L'}{L} = 0 \quad (45.2-42)$$

亦即  $L' = 0$ , 可见使  $\ln L$  取得极大值时也会使  $L$  取得极大值。

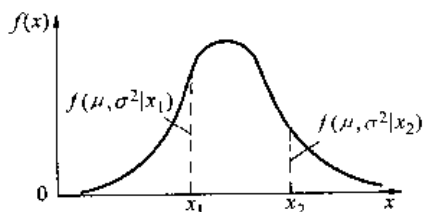


图 45.2-3 极大似然性原理

上述概念不难推广到  $n$  元正态变量的情况。现有

一容量为  $n$  的子样  $(x_1, x_2, \dots, x_n)$ , 为了由此估计母体参数  $\mu$  和  $\sigma^2$ , 首先写出似然性函数

$$\begin{aligned} L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) \\ &= f(\mu, \sigma^2 | x_1) f(\mu, \sigma^2 | x_2) \dots \\ &\quad f(\mu, \sigma^2 | x_n) \\ &= \prod_{i=1}^n f(\mu, \sigma^2 | x_i) = \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(x_i - \mu)^2}{2\sigma^2} \right] \end{aligned}$$

上式还可写成

$$\begin{aligned} L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = \\ \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \end{aligned} \quad (45.2-43)$$

将式(45.2-43)取自然对数

$$\begin{aligned} L(\mu, \sigma^2 | x_1, x_2, \dots, x_n) = \\ -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned} \quad (45.2-44)$$

分别求出上式对  $\mu$  和  $\sigma^2$  的偏导数,并令其等于

零。求偏导数时,注意到  $\frac{\partial \ln \sigma^2}{\partial \sigma^2} = \frac{1}{\sigma^2}$ , 则有

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad (45.2-45)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \quad (45.2-46)$$

式(45.2-45)还可写成

$$\sum_{i=1}^n x_i - n\mu = 0$$

由此可求得母体平均值估计量

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (45.2-47)$$

再用  $\bar{x}$  替换式(45.2-46)中的  $\mu$ , 则得到母体方差估计量

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (45.2-48)$$

用极大似然法求出的母体参数估计量称作“极大似然性估计量”,它可能是无偏估计量,也可能是有偏估计量。以上求出的母体方差估计量是有偏估计量,只有当  $n$  很大时,才与无偏估计量(子样方差  $s^2$ )接近。

## 第 3 章 机械零件可靠性设计

### 1 机械零件可靠性设计的特点、内容与 方法

#### 1.1 机械零件可靠性设计的特点

传统机械零件设计采用确定性的材料性能参数、工作许用应力和安全系数,不考虑参数的分散性。可靠性设计(或称概率设计)则把设计指标及有关参数作为随机变量处理,在给定的载荷条件下,应用统计与概率分析法,设计出安全、合理的零件或结构。可靠性设计的基本特点是将应力和强度作为随机变量处理,应用概率和统计方法进行分析、计算,定量地评价产品的失效率和可靠度。

#### 1.2 机械零件可靠性设计的基本内容及步骤

- 1) 提出设计任务,规定详细的可靠度指标;
- 2) 确定有关的设计变量及参数;
- 3) 确定零件的应力分布;
- 4) 确定零件强度计算公式;
- 5) 确定零件的强度分布;
- 6) 应用应力-强度干涉模型,采用合适的计算方法,计算零件的可靠度;
- 7) 确定零件可靠度的置信度区间。

### 2 应力-强度干涉模型与可靠度计算方法

#### 2.1 应力-强度干涉模型

令  $f_1(x_1)$  为应力分布的概率密度函数,  $f_2(x_2)$  为强度分布的概率密度函数,两者一般存在干涉区,如图 45.3-1 所示。

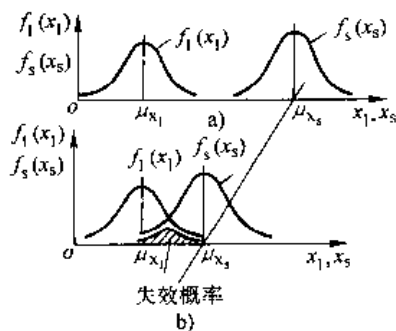


图 45.3-1 应力-强度干涉模型

则可靠度为  $R = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{x_2} [f_1(x_1) dx_1] f_2(x_2) dx_2 \right]$

或  $R = \int_{-\infty}^{+\infty} \left[ \int_{x_1}^{+\infty} [f_2(x_2) dx_2] f_1(x_1) dx_1 \right]$

(45.3-1)

#### 2.2 应力分布的确定

机械零件所受的工作应力  $S$  与其承受的载荷、温度、几何尺寸、物理特性、时间等参数有关。确定应力分布的步骤为:

- 1) 确定零件的失效模式及其判据;
- 2) 对零件进行应力分析;
- 3) 确定失效应力、应力修正系数及有关设计参数的分布;
- 4) 确定可靠性计算用应力分布。

##### 2.2.1 零件尺寸误差

零件上作用的常见载荷有拉、压、弯、扭。设载荷是等幅循环载荷,载荷产生的应力也是等幅应力。由于每个零件的尺寸在批量生产中会各有不同,在公差范围内合格的零件,其尺寸必须作为随机变量来处理。再加上载荷也是随机变量,危险点的应力自然也是随机变量,一般假设它服从正态分布。

通常,机械加工中的容许尺寸偏差是用公差来表示的。如果现有的与尺寸的变异性有关的惟一数据是容许偏差  $\pm \Delta x$ ,常常可以利用它来估计标准差  $\sigma$ 。当预期的数据能够集中在  $\bar{x} \pm \Delta x$  的界限内,这个界限便可用来确定一个大子样的幅度。通常,标准差的近似值为

$$S_x \approx \frac{(\bar{x} + \Delta x) - (\bar{x} - \Delta x)}{6} = \frac{\Delta x}{3} \quad (45.3-2)$$

一般认为尺寸服从正态分布,  $S_x$  与  $\Delta x$  的关系如图 45.3-2 所示。

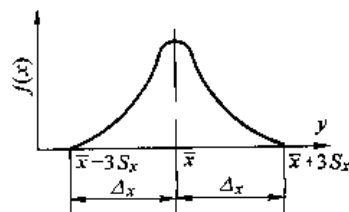


图 45.3-2  $S_x$  与  $\Delta x$  的关系

例 已知宽度为 1000~1200mm 的冷轧碳钢板的名义厚度为  $t = 4.60 \sim 4.80\text{mm}$ , 容许偏差为  $\pm 0.250\text{mm}$ , 问钢板厚度的标准差为多大?

解 由式(45.3-2)可得

$$\sigma_t \approx \frac{\Delta t}{3} = \frac{0.250}{3} = 0.083\text{mm}$$

式(45.3-2)不仅可以用于确定几何尺寸的标准差,也可用于确定载荷的标准差。

显然,当误差对称于公称尺寸时,可取公称尺寸为均值  $\bar{x}$ ,取  $\Delta x/3$  为标准差  $S_x$ 。若误差不对称于公称尺寸,可根据公称尺寸和误差先求出最大值  $x_{\max}$  和最小值  $x_{\min}$ ,然后将均值和标准差分别取为

$$\bar{x} = \frac{x_{\max} + x_{\min}}{2} \quad (45.3-3)$$

表 45.3-1 正态分布随机变量的代数运算公式

序号	$z$ 的运算	均 值 $\mu_z$	标 准 离 差 $\sigma_z$
1	$z=c$	$c$	0
2	$z=cx$	$c\mu_x$	$c\sigma_x$
3	$z=cx \pm d$	$c\mu_x \pm d$	$c\sigma_x$
4	$z=x+y$	$\mu_x + \mu_y$	$(\sigma_x^2 + \sigma_y^2)^{\frac{1}{2}}$ 或 $(\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y)^{\frac{1}{2}}$
5	$z=x-y$	$\mu_x - \mu_y$	$(\sigma_x^2 + \sigma_y^2)^{\frac{1}{2}}$ 或 $(\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y)^{\frac{1}{2}}$
6	$z=xy$	$\mu_x\mu_y$ 或 $\mu_x\mu_y + \rho\sigma_x\sigma_y$	$(\mu_x^2\sigma_y^2 - \mu_y^2\sigma_x^2 + 2\rho\mu_x\mu_y\sigma_x\sigma_y)^{\frac{1}{2}}$ 或 $[(\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2)(1+\rho^2)]^{\frac{1}{2}}$
7	$z=\frac{x}{y}$	$\frac{\mu_x}{\mu_y}$ 或 $\frac{\mu_x}{\mu_y} + \frac{\mu_x\sigma_y^2}{\mu_y^3} \left( \frac{\sigma_y}{\mu_y} - \rho \frac{\sigma_x}{\mu_x} \right)$	$\frac{1}{\mu_y} \left( \frac{\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2}{\mu_y^2 + \sigma_y^2} \right)^{\frac{1}{2}}$ 或 $\frac{1}{\mu_y^2} (\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2)^{\frac{1}{2}}$ 或 $\frac{\mu_x}{\mu_y} \left( \frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - 2\rho \frac{\sigma_x\sigma_y}{\mu_x\mu_y} \right)^{\frac{1}{2}}$
8	$z=x^2$	$\mu_x^2$ 或 $\mu_x^2 + \sigma_x^2$	$2\mu_x\sigma_x$ 或 $(4\mu_x^2\sigma_x^2 + 2\sigma_x^4)^{\frac{1}{2}}$
9	$z=x^3$	$\mu_x^3$ 或 $\mu_x^3 + 3\sigma_x^2\mu_x$	$3\mu_x^2\sigma_x$ 或 $(3\sigma_x^4 + 8\sigma_x^2\mu_x^2 + 5\sigma_x^2\mu_x^4)^{\frac{1}{2}}$
10	$z=x^n$	$\mu_x^n$	$n\mu_x^{n-1}\sigma_x$
11	$z=x^{\frac{1}{2}}$	$\left( \frac{1}{2} \sqrt{4\mu_x^2 - 2\sigma_x^2} \right)^{\frac{1}{2}}$	$\left( \mu_x - \frac{1}{2} \sqrt{4\mu_x^2 - 2\sigma_x^2} \right)^{\frac{1}{2}}$
12	$z=(x^2+y^2)^{\frac{1}{2}}$	$(\mu_x^2 + \mu_y^2)^{\frac{1}{2}}$	$\left( \frac{\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2}{\mu_x^2 + \mu_y^2} \right)^{\frac{1}{2}}$

$$S_x = \frac{x_{\max} + x_{\min}}{6} \quad (45.3-4)$$

2.2.2 应力分布参数的代数运算

如果影响零件工作应力  $S$  的参数  $X_1 \sim (\mu_1, \sigma_1^2)$ ,  $X_2 \sim (\mu_2, \sigma_2^2), \dots, X_n \sim (\mu_n, \sigma_n^2)$  均为正态随机变量,则可以根据这些参数与应力的函数关系,把它们综合为仅含单一随机变量  $Z$  的应力函数  $S(Z) = f(X_1, X_2, \dots, X_n)$ , 并确定其分布。

如果各随机变量的变异系数都小于 0.1, 即  $\sigma_i/\mu_i < 0.1$ , 且满足随机变量的多重性要求,则由中心极限定理可知,这个应力函数近似于正态分布,正态分布随机变量的代数运算公式列于表 45.3-1。

2.3 强度分布

确定强度分布的步骤为

- 1) 确定强度判据;
- 2) 确定名义强度分布;
- 3) 修正名义强度分布。

需要注意的是,材料性能手册给出的材料数据,多是用标准试样得到的材料的名义强度。如果将名义强度转化成零件的实际强度,必须进行修正。由此,引入了应力集中系数、尺寸系数和表面系数,这些系数本身也是随机变量,一般服从正态分布。

在材料性能手册中,可以查到各种材料在不同载荷类型作用下的应力集中系数。在此仅给出钢件尺寸

系数的统计参数(见表 45.3-2),钢件表面加工系数的统计参数(见表 45.3-3 和表 45.3-4)。

表 45.3-2 钢件尺寸系数  $\epsilon$  的统计参数

钢种	尺寸 $d$ /mm	试样数 $n$	$\epsilon$ 的统计参数		
			均值 $\epsilon$	标准差 $s_\epsilon$	变异系数 $\nu = s_\epsilon/\bar{\epsilon}$
碳素钢	30~150	8	0.8562	0.08895	0.10388
	150~250	8	0.8025	0.04773	0.05948
	250~350	9	0.7911	0.03444	0.04353
	350 以上	14	0.73	0.04188	0.05737
合金钢	30~150	11	0.79	0.0690	0.08734
	150~250	12	0.7667	0.07487	0.09765
	250~350	5	0.678	0.06834	0.10079
	350 以上	22	0.6718	0.07202	0.10720

表 45.3-3 表面加工系数的均值  $\bar{\beta}$  及标准差  $s_{\beta}$

钢种	锻造		粗车 $R_a=12.5\mu\text{m}$		精车 $R_a=3.2\mu\text{m}$		磨削 $R_a=0.4\mu\text{m}$		抛光 $R_a=0.1\mu\text{m}$	
	$\bar{\beta}$	$s_{\beta}$	$\bar{\beta}$	$s_{\beta}$	$\bar{\beta}$	$s_{\beta}$	$\bar{\beta}$	$s_{\beta}$	$\bar{\beta}$	$s_{\beta}$
35	0.8795	0.0292	0.9816	0.0166	0.9868	0.0255	1	0.0169	1.0112	0.0228
45	0.6386	0.0160	0.9668	0.0205	0.9873	0.0224	1	0.0221	1.0079	0.0241
16Mn	0.6061	0.0148	0.8367	0.0193	0.9104	0.0180	1	0.0190	1.0007	0.0202
40Cr	0.5353	0.0209	0.8479	0.0505	0.9011	0.0355	1	0.0499	1.0210	0.0401
60Si2Mn	0.4560	0.0173	0.7622	0.0307	0.8661	0.0346	1	0.0353	1.0143	0.0331

表 45.3-4 表面加工系数的变异系数  $\nu_{\beta}(=s_{\beta}/\bar{\beta})$

钢种	表面加工	锻造	粗车 $R_a=12.5\mu\text{m}$	精车 $R_a=3.2\mu\text{m}$	磨削 $R_a=0.4\mu\text{m}$	抛光 $R_a=0.1\mu\text{m}$	不同加工平均
35		0.033	0.017	0.026	0.017	0.023	0.023
45		0.025	0.021	0.023	0.022	0.024	0.023
16Mn		0.024	0.023	0.020	0.019	0.020	0.021
40Cr		0.039	0.060	0.039	0.050	0.039	0.045
60Si2Mn		0.038	0.040	0.040	0.035	0.033	0.037

2.4 可靠度计算方法

根据应力-强度干涉模型,求机械零件可靠度的主要方法有公式法、数值积分法、图解法。此外,还有根据极限状态方程的验算点法和蒙特卡洛模拟法等。

2.4.1 应力-强度干涉模型求可靠度的一般公式

根据应力和强度分布函数计算的简便性,可选择下面二式之一计算可靠度

$$R = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{x_s} f_1(x_1) dx_1 \right] f_2(x) dx$$

$$= \int_{-\infty}^{+\infty} F_1(x) f_2(x) dx \quad (45.3-5)$$

$$R = \int_{-\infty}^{+\infty} \left[ \int_{x_1}^{+\infty} f_2(x_2) dx_2 \right] f_1(x_1) dx_1$$

$$= 1 - \int_{-\infty}^{+\infty} F_2(x) f_1(x) dx \quad (45.3-6)$$

式中  $F_1(x)$  - 应力分布函数;  
 $F_2(x)$  - 强度分布函数;  
 $f_1(x)$  - 应力概率密度;  
 $f_2(x)$  - 强度概率密度。

几种典型分布情况下根据应力-强度干涉模型求可靠度公式见表 45.3-5。

表 45.3-5 几种典型分布求可靠度公式

序号	应力	强度	可靠度公式
1	正态 $N(\bar{x}_1, s_1^2)$	正态 $N(\bar{x}_2, s_2^2)$	$R = \int_{x_0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1 - \Phi(z_p) - \Phi(z_R)$ $z_R = \frac{\bar{x}_2 - \bar{x}_1}{(s_1^2 + s_2^2)^{\frac{1}{2}}}$ 正态变量 $z_R$ 称为联结系数
2	对数正态 $\ln(\mu, \sigma^2)$ 或 $\lg(\mu, \sigma^2)$	对数正态 $\ln(\mu, \sigma^2)$ 或 $\lg(\mu, \sigma^2)$	$z_R = \frac{\mu_2 - \mu_1}{(\sigma_2^2 - \sigma_1^2)^{\frac{1}{2}}} \approx \frac{\ln \bar{x}_2 - \ln \bar{x}_1}{(V_2^2 - V_1^2)^{\frac{1}{2}}}$ , $V_i = \frac{s_i^2}{\bar{x}_i^2}$ , $\sigma^2 = \ln(V_i^2 + 1)$ , $\mu = \ln \bar{x} - \frac{\sigma^2}{2}$ 或 $z_R = \frac{\mu_2 - \mu_1}{(\sigma_2^2 + \sigma_1^2)^{\frac{1}{2}}} \approx \frac{2.303(\lg \bar{x}_2 - \lg \bar{x}_1)}{(V_2^2 - V_1^2)^{\frac{1}{2}}}$ , $\sigma^2 = 0.4343 \lg(V_i^2 + 1)$ , $\mu = \lg \bar{x} - 1.151 \sigma^2$
3	指数 $e(\lambda)$	指数 $e(\lambda)$	$R = \frac{\lambda_1}{\lambda_1 + \lambda_2}$
4	正态 $N(\bar{x}_1, s_1^2)$	指数 $e(\lambda)$	$R = e^{-\frac{1}{2}(2\bar{x}_1\lambda - \lambda^2 s_1^2)}$
5	指数 $e(\lambda_1)$	正态 $N(\bar{x}_2, s_2^2)$	$R = 1 - e^{-\frac{1}{2}(2\bar{x}_2\lambda_1 - \lambda_1^2 s_2^2)}$
6	指数 $e(\lambda)$	$\Gamma$ $\Gamma(\alpha, \beta)$	$R = 1 - \left( \frac{\beta_2}{\beta_2 - \lambda_1} \right)^{\alpha}$
7	$\Gamma$ $\Gamma(\alpha, \beta)$	指数 $e(\lambda)$	$R = \left( \frac{\beta_1}{\beta_1 - \lambda_2} \right)^{\alpha}$
8	$\Gamma$ $\Gamma(\alpha, \beta)$	正态 $N(\bar{x}_2, s_2^2)$	$R = 1 - (1 - \bar{x}_2\beta_1 - s_2^2\beta_1^2) e^{-\frac{1}{2}(s_2^2\beta_1^2 - 2\bar{x}_2\beta_1)}$
9	瑞利 $R(\mu)$	正态 $N(\bar{x}_2, s_2^2)$	$R = 1 - \frac{\mu}{(\mu^2 + s_2^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left( \frac{\bar{x}_2^2}{\mu^2 + s_2^2} \right)}$

注:  $\Phi(\cdot)$  为标准正态分布函数查《机械设计手册》第2版表 13.3-14(或第2篇)。

**例** 钢制拉杆, 工作应力  $x_1 = \sigma \sim N(400, 25^2)$  MPa, 屈服强度  $x_2 = \sigma_s \sim N(500, 50^2)$  MPa, 求不发生屈服失效的可靠度。

**解** 按表 45.3-5 中序号 1 列出的公式

$$z_R = \frac{\bar{x}_2 - \bar{x}_1}{(s_1^2 + s_2^2)^{\frac{1}{2}}} = \frac{500 - 400}{(25^2 + 50^2)^{\frac{1}{2}}} = 1.789$$

查表 45.2-4 或计算可得

$$R = \phi(z_R) = \phi(1.789) = 0.96318$$

### 2.4.2 数值积分法求可靠度

有些应力和强度的分布用式(45.3-5)或(45.3-6)难以积分, 没有象表 45.3-5 中那样简单的结论式可用, 这时可用数值积分, 例如用梯形公式或辛普森公式等, 并可利用计算机进行运算。下面仅按式(45.3-5)或(45.3-6)作数值积分计算, 这时可导得下列近似式(参看图 45.3-3)。

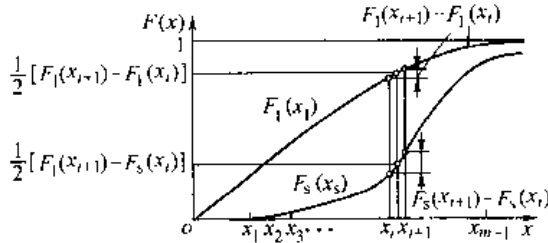


图 45.3-3 数值积分  $x$  轴的划分

$$R \approx \frac{1}{2} \sum_{i=0}^{m-1} [F_1(x_{i+1}) + F_1(x_i)] \times [F_2(x_{i+1}) + F_2(x_i)] \quad (45.3-7)$$

或 
$$R \approx 1 - \frac{1}{2} \sum_{i=0}^{m-1} [F_1(x_{i+1}) - F_1(x_i)] \times [F_2(x_{i+1}) - F_2(x_i)] \quad (45.3-8)$$

由于  $x$  轴为  $0 \sim \infty$ , 数值计算只在区间  $(x_1, x_{m-1})$  进行细算,  $x_1$  为  $F_1(x)$  和  $F_2(x)$  接近于 1 中较大的  $x$  值,  $x_{m-1}$  为  $F_1(x)$  和  $F_2(x)$  接近于 1 中较小的  $x$  值。区间初选如不合适, 可在计算中适当增减。区间取得越大, 在区间内划分越细, 则结果越精, 但计算量越大。一般宜列表计算。

**例** 某构件应力为正态分布,  $\bar{x}_1 = 100$  MPa,  $s_1 = 10$  MPa; 强度为威布尔分布,  $k = 3, a = 50$  MPa,  $b = 200$  MPa, 求可靠度。

**解** 先选区间  $(x_1, x_{m-1})$ , 选择  $x_1 = a = 50$  MPa 和  $x_1 = \bar{x}_1 - 3s_1 = 100 - 3 \times 10$  MPa = 70 MPa 中的大值, 故选定  $x_1 = 70$  MPa,  $x_{m-1} = \bar{x}_1 + 3s_1 = 100 + 3 \times 10$  MPa = 130 MPa。在  $(x_1, x_{m-1})$  区间内。

每点递增 5 MPa。  $F_1(x_i) = \phi(z_i)$ , 按  $z_i = \frac{x_i - \bar{x}_1}{s_1}$  查正态分布表(表 45.2-4)。  $F_2(x_i) = 1 - e^{-\left(\frac{x_i - a}{b}\right)^k}$  算得。按式(45.3-7)列表计算(见表 45.3-6), 故得

$$R = \frac{1}{2} \times 1.963849 = 0.981925$$

表 45.3-6 计算结果

$i$	$x_i$	$z_i = \frac{x_i - 100}{10}$	$F_1(x_i) = \Phi(z_i)$	$F_1(x_{i+1}) + F_1(x_i)$	$F_2(x_i) = 1 - e^{-\left[\left(\frac{x_i - 50}{200}\right)^3\right]}$	$F_2(x_{i+1}) - F_2(x_i)$	$[F_1(x_{i+1}) + F_1(x_i)] [F_2(x_{i+1}) - F_2(x_i)]$
0	0	-10	0.000	0.001	0	0.001	0.000001
1	70	-3	0.001	0.007	0.001	0.001	0.000007
2	75	-2.5	0.006	0.029	0.002	0.001	0.000029
3	80	-2	0.023	0.090	0.003	0.002	0.000180
4	85	-1.5	0.067	0.226	0.005	0.003	0.000678
5	90	-1	0.159	0.460	0.008	0.003	0.001380
6	95	-0.5	0.301	0.801	0.011	0.005	0.004005
7	100	0.0	0.500	1.191	0.016	0.005	0.005955
8	105	0.5	0.691	1.532	0.021	0.006	0.009192
9	110	1.0	0.841	1.774	0.027	0.007	0.012418
10	115	1.5	0.933	1.910	0.034	0.008	0.015280
11	120	2.0	0.977	1.971	0.042	0.009	0.017739
12	125	2.5	0.994	1.993	0.051	0.011	0.021923
13	130	3.0	0.999	1.999	0.062	0.938	1.875062
14	$\infty$	$\infty$	1.000		1.000		

$\Sigma 1.963849$

### 2.4.3 梅林变换法求可靠度

梅林(Mellin)变换法, 可适用于任何两种不同分布的组合, 是较实用的计算可靠度的图解法。可靠度的通用式(45.3-6)

$$R = \int_{-\infty}^{+\infty} f_1(x_1) \left[ \int_{-\infty}^{+\infty} f_2(x_2) dx_2 \right] dx_1$$

令 
$$G = \int_{-\infty}^{+\infty} f_2(x_2) dx_2 \quad (45.3-9)$$

$$H = \int_1^x f_1(x_1) dx_1 \quad (45.3-10)$$

则式(45.3-6)变成

$$R = \int_c^1 G dH \quad (45.3-11)$$

式中  $dH=f(x)dx$ 。

由于积分变量的改变,所以式(45.3-11)的积分限也随之改变。由累积分布函数的性质可知, $G$ 和 $H$ 值变化的最大范围是从0到1。图45.3-4表示给定值,可由式(45.3-9)和式(45.3-10)求出不同的 $G$ 和 $H$ :当 $x_1=-\infty$ 时, $G=H=1.0$ ;当 $x_1=\infty$ 时, $G=H=0$ 。所以,可以在区间 $[0,1]$ 内设计一条曲线 $G=f(H)$ ,使该曲线下的面积等于式(45.3-11)表达的可靠度 $R$ ,如图45.3-5所示。

对于不同的 $x$ 值,可由式(45.3-9)和式(45.3-10)求出不同的 $G$ 和 $H$ 值,把它们对应地画在图45.3-5上,便可绘出 $G=f(H)$ 曲线。测量出曲线下方的面积,即可得到所求的可靠度。

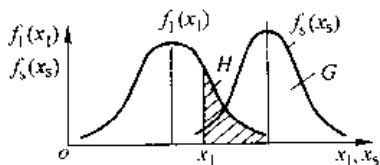


图 45.3-4 概率密度函数曲线与  $G$ 、 $H$  的关系

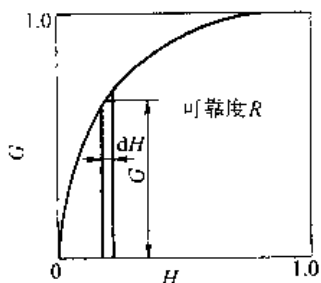


图 45.3-5 求  $R$  示意图

例 图 45.3-6 所示的拉力板,板厚为  $8.68 \pm 0.09\text{mm}$  设应力为威布尔分布,强度为正态分布,求拉力板的可靠度。

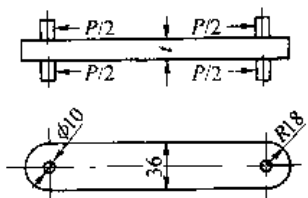


图 45.3-6 拉力板

解 1) 设 40CrNiMoA 钢的拉力板  $N=10^6$  时的疲劳极限线图已知,如图 45.3-7 所示。并已知应力比  $r=0.2857$ 。该应力比  $r$  直线与拉力板疲劳极限线图中

的均值曲线相交于  $M$  点,与  $R=0.99999$  的曲线相交于  $N$  点。量得  $M$  点的坐标为  $255\text{MPa}$  和  $138\text{MPa}$ ;  $N$  点的坐标为  $195\text{MPa}$  和  $100\text{MPa}$ ,则  $M$  点距坐标原点  $O$  的距离等于强度正态分布均值  $\bar{x}$ 。

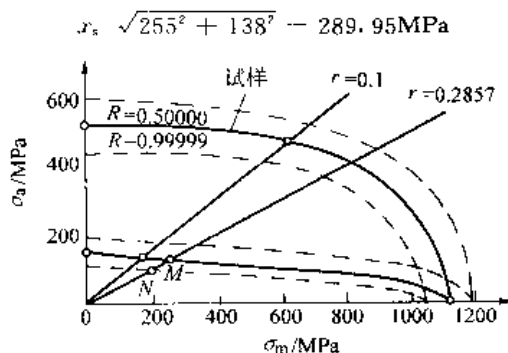


图 45.3-7 拉力板  $N=10^6$  时的疲劳极限线图

$N$  点高原点  $O$  的距离 =  $\sqrt{195^2 + 100^2} = 219.146\text{MPa}$ 。

由表 45.2-4 查得当  $R=0.99999$  时的可靠度系数  $z=4.265$ ,所以拉力板疲劳极限的标准差  $s$  为

$$s = \frac{289.95 - 219.146}{4.265} \text{MPa} = 16.6\text{MPa}$$

即拉力板的强度(正态分布)为  $(289.95, 16.6)$  MPa

2) 求可靠度  $R$  的通用式

$$R = \int_{-\infty}^{\infty} f_1(x_1) \left[ \int_{x_1}^{\infty} f_2(x_2) dx_2 \right] dx_1$$

当  $f_1(x_1)$  为威布尔分布的特定情况下,其积分下限  $-\infty$  应改为位置参数  $a$ , 即有

$$R = \int_a^{\infty} f_1(x_1) \left[ \int_{x_1}^{\infty} f_2(x_2) dx_2 \right] dx_1 \quad (a)$$

威布尔分布概率密度函数

$$f_1(x_1) = \frac{k}{b} \left( \frac{x_1 - a}{b} \right)^{k-1} e^{-\left( \frac{x_1 - a}{b} \right)^k} \quad (b)$$

因  $R = \int_a^{\infty} f_1(x_1) dx_1 = 1$ , 进行下面的积分,并考虑式(b)的积分  $F_1(x_1) = 1 - e^{-\left( \frac{x_1 - a}{b} \right)^k}$ , 有

$$\begin{aligned} & \int_a^x \frac{k}{b} \left( \frac{x_1 - a}{b} \right)^{k-1} e^{-\left( \frac{x_1 - a}{b} \right)^k} dx_1 \\ &= \int_a^{\infty} \frac{k}{b} \left( \frac{x_1 - a}{b} \right)^{k-1} e^{-\left( \frac{x_1 - a}{b} \right)^k} dx_1 \\ & - \int_x^{\infty} \frac{k}{b} \left( \frac{x_1 - a}{b} \right)^{k-1} e^{-\left( \frac{x_1 - a}{b} \right)^k} dx_1 \\ &= 1 - f_1(x_1) dx_1 \\ &= 1 - \left[ 1 - e^{-\left( \frac{x_1 - a}{b} \right)^k} dx_1 \right] \end{aligned}$$

考虑式(45.3-10)的关系,上式为  $H$ , 即

$$H = e^{-\left( \frac{x_1 - a}{b} \right)^k} \quad (c)$$



将应力的实测数据在威布尔概率坐标纸上进行分析,得式(45.3-10)中的各参数值为: $k=1.25, b=85\text{MPa}, \alpha=40\text{MPa}$ 代入式(c),有

$$H = e^{-\left(\frac{x-40}{85}\right)^{1.25}} \quad (d)$$

3) 强度为正态分布,一般经过变换。可以将非标准正态分布函数,化为标准正态分布函数,即将均值移至零。为此,令随机变量

$$u = \frac{x - \mu}{\sigma}$$

$u$  称为标准正态偏量,有

$$u = \frac{x - \bar{x}}{s} = \frac{x - 289.95}{16.6} \quad (e)$$

因为  $u$  在数值上等于可靠度系数  $L_R$ ,即可靠度  $R$  可以算出,计算步骤如表 45.3-7。将表中的  $(F_i, G_i)$  描点,过  $F_i$  点作垂线,将分成的小梯形求和,得相当于式(45.3-11)的积分,即图 45.3-8 右侧的面积。

$$R = \frac{1}{2} \left[ \sum_{i=1}^{15} (G_i + G_{i+1})(H_i - H_{i+1}) \right] = 0.9772$$

即拉力板的可靠度为  $R=0.9772$ 。

表 45.3-7 梅林变换法求可靠度例题计算结果

序号	$x_i, y_i$ /MPa	①	②	③	④	⑤	⑥
		$H_i$ [式(d)]	$z_i$ [式(e)]	$G_i$	$(H_i - H_{i+1})$	$\frac{1}{2}(G_i + G_{i+1})$	④×⑤
1	40	1.000000	-15.057	~1.000000	0.1511523	~1.000000	0.1511523
2	60	0.8488477	-13.852	~1.000000	0.1716308	~1.000000	0.1716308
3	80	0.6772160	-12.647	~1.000000	0.1536111	~1.000000	0.1536111
4	100	0.5236055	-11.443	~1.000000	0.1278741	~1.000000	0.1278741
5	120	0.3957314	-10.238	~1.000000	0.1020485	~1.000000	0.1020485
6	140	0.2936829	-9.0331	~1.000000	0.0790602	~1.000000	0.0790602
7	160	0.2146227	7.8283	~1.000000	0.0598646	~1.000000	0.0598646
8	180	0.1547581	-6.6235	~1.000000	0.0441896	~1.000000	0.0441896
9	200	0.1102685	-5.4187	~1.000000	0.0325425	0.9999934	0.0325422
10	220	0.077726	4.2139	0.999987	0.0234766	0.9993355	0.023461
11	240	0.0542492	-3.0090	0.99869	0.0167293	0.981495	0.0164197
12	260	0.0375201	-1.8042	0.96130	0.0117897	0.845025	0.0099626
13	280	0.0257304	-0.5992	0.72575	0.0082247	0.499175	0.0041056
14	300	0.0175057	-0.6054	0.27260	0.0056846	0.1539	0.008749
15	320	0.0118211	-1.8102	0.03520	0.00591055	0.0176	0.000104
16		0.0000000		0.00000			
总计							0.977201

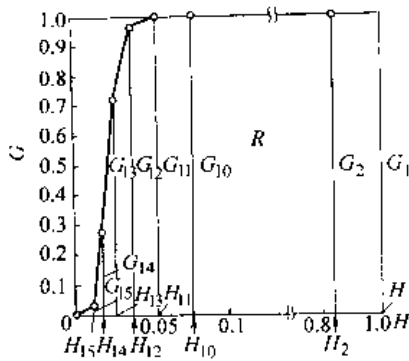


图 45.3-8 梅林变换法求可靠度例题

分布中随机抽取一个应力值,再与从强度分布随机抽取一个强度值进行比较,如果应力大于强度,则零件失效,反之,零件安全。由此可见,每一次随机模拟相当于对一个随机抽取的零件进行一次试验。

设模拟次数为  $N$ ,失效数为  $F$ ,则零件的失效概率近似值等于  $F/N$ ,可靠度  $R=1-\frac{F}{N}$ 。模拟次数  $N$  越大,则模拟精度越高。要获得可靠的模拟结果,往往需要进行千次以上的模拟。由于模拟次数多,所以,用蒙特卡罗法进行模拟,一般由计算机完成。

在计算机上进行零件可靠度的模拟,其大致过程如下:

- 1) 确定随机模拟次数  $N$  及计算可靠度的公式; $P(x_1 - x_2) > 0$ ;
- 2) 输入原始资料 原始资料包括:零件工作应力  $x_1$  及强度  $x_2$  的概率密度函数,影响应力和强度的各独立随机变量的参数等;
- 3) 给出各独立随机变量的分布规律;
- 4) 产生符合上述给定分布规律的随机数组,每个

### 2.4.4 蒙特卡罗法

#### (1) 蒙特卡罗法求解可靠度的原理

蒙特卡罗法是通过随机变量的统计试验或随机模拟,求解工程技术问题近似解的方法。因此,也称为统计试验法或随机模拟法。

应用蒙特卡罗法求解可靠度,必须知道应力和强度分布函数的类型及参数。这种方法的实质是从应力

随机数组中各有  $N$  个随机数；

5) 计算零件的工作应力 在影响零件工作应力的随机数组中,抽取一组随机数代入应力公式,计算零件的工作应力  $x_1$ ;

6) 计算零件的强度 在影响零件强度的随机数组中,抽取一组随机数代入强度公式,计算零件的强度值  $x_2$ ;

7) 比较应力及强度 若  $x_1 - x_2 > 0$ ,则零件失效;反之,零件安全。如此重复进行  $N$  次,可以得到  $N$  次模拟时零件的失效数  $F$ ;

8) 计算零件可靠度的模拟值  $R = 1 - \frac{F}{N}$  模拟零件可靠度的计算框图如图 45.3-9。

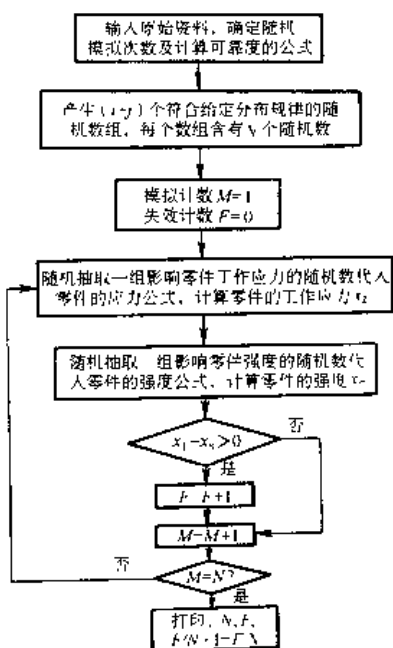


图 45.3-9 蒙特卡罗法可靠度的计算框图

从零件可靠度随机模拟的过程可以看出,蒙特卡罗法的主要优点是可以确定复杂随机变量的概率分布和数字特征。此外,对于用于干涉理论解析法难以处理的分布,也可用蒙特卡罗法求解。

(2) 随机分布的抽样

蒙特卡罗法中重要的一步是生成随机数。产生随机变量的基础是产生  $[0, 1]$  区间上均匀分布的随机数(即随机数发生器)。其他各类分布如正态分布、对数正态分布、泊松分布等,都是通过  $[0, 1]$  区间上的均匀分布进行变换得到的,所以随机数的产生至关重要。

1) 随机数的产生方法 目前计算机仿真中采用的随机数发生器都是按一定的数学函数的递推算法得到的,因此所产生的数严格讲,不是概率意义上真正的随机数,被称为伪随机数。但如果算法选择合适,所得

到的伪随机数序列周期足够长,而且能满足均匀性、独立性要求,将这种伪随机数用于仿真是可行的。产生随机数的常用方法有同余法、混合同余法、组合法等。

a) 运用乘同余法产生均匀分布随机数的迭代公式为

$$x_i = a \cdot x_{i-1} \pmod{m} \quad (i = 1, 2, 3 \dots) \tag{45.3-12}$$

式中  $x_i$  为第  $i$  个随机数,  $a$  为乘子,  $m$  为模,两者均为正整数,  $\pmod{m}$  表示除以  $m$  后取余数,当  $x_0$  给定时用上式递推可以得到一组数的序列,然后将该序列的每个值除以模  $m$ ,便可以得到  $[0, 1]$  间上均匀分布的随机数。一般来说以  $2^t$  为模的乘同余随机数发生器统计性质不好,很少使用。为了提高乘同余法的可用性,较为理想的方法是,  $m$  取素数(只能被 1 和自身整除的数)。  $a$  选取条件是:选择  $a$  使  $a^{m-1} - 1$  为能被  $m$  整除的最小整数,此时得到的随机数具有最大周期。

b) 混合同余法产生随机数的迭代公式为

$$x_i = (a \cdot x_{i-1} + c) \pmod{m} \quad (i = 1, 2, 3 \dots) \tag{45.3-13}$$

$$r_i = \frac{x_i}{m}$$

式中  $x_i$  ——第  $i$  个随机数;

$a$  ——乘子;

$c$  ——增量;

$m$  ——模;

$x_0$  ——随机数源或种子。

如果  $m$  足够大,且发生器具有满足周期,可以用  $r_i$  代替均匀分布的随机变量。加大  $m$  可以提高伪随机数的均匀性。在实际应用中,一般取  $m = 2^t$ ,  $c$  为奇数,  $a - 1$  能被 4 整除,将得到满周期随机数序列,其中  $x_0$  可以去任意整数,如果需要多次运行模拟程序,而每次运行需采用不同的均匀随机数时,  $x_0$  可以取当前时间值。

c) 组合发生器的基本思想是将两组独立的随机数序列组合起来,形成新的随机数序列以提高线性同余发生器的性能、减少随机数序列的自相关,一般来说两个性能不太好的随机数发生器组合后能得到统计特性较好的随机数序列。乘同余法和混合同余法统称为线性同余法,方法简单,程序容易实现,因而应用广泛。

2) 伪随机数序列的检验 由于数学方法产生的随机数有时并非均匀分布,而且序列间具有一定的相关性,在使用前应进行随机数发生器的均匀性和独立性检验。

a) 均匀性检验 最常用的均匀性检验方法是频率检验法,即将随机数的取值范围  $[0, 1]$  分成个互不重叠的等长的  $K$  子区间,计算由随机数发生器产生的  $N$

个随机数落在一个子区间的概率  $P = n_j/N$ , ( $n_j$  为落在第  $j$  个子区间的随机数的个数), 均匀分布的理论概率为  $P = 1/K$ , 实际频率与理论频率的偏差为

$$\chi^2 = \sum_{j=1}^K \left( \frac{n_j}{N} - \frac{1}{K} \right)^2 \quad (45.3-14)$$

$\chi^2$  的大小反映了随机数的均匀程度。当  $K$  足够大时,  $\chi^2$  近似服从自由度为  $K-1$  的卡方分布。

b) 独立性检验 独立性检验的常用的方法是计算序列相邻一定间隔的随机数之间的相关系数, 判断其相关程度。对于由随机数发生器产生的  $N$  个随机数  $r_i$ , 前后相隔  $j$  个数的两个随机数  $r_i$  和  $r_{i+j}$  的相关系数计算公式为

$$\hat{\rho}_j = \frac{1}{S^2} \left[ \frac{1}{N-j} \cdot \sum_{i=1}^{N-j} r_i \cdot r_{i+j} - \bar{r}^2 \right] \quad (45.3-15)$$

其中  $S^2$  为随机数的方差。取统计量  $\mu = \hat{\rho} \cdot \sqrt{N-j}$ , 当  $N$  充分大时,  $\mu$  渐近服从标准正态分布  $N(0, 1)$ 。因此通过  $\mu$  来检验随机数序列的独立性。

3) 抽样方法 在含有随机变量的模型中, 随机变量是通过随机抽样的方法进行模拟的。常用的产生给定分布的随机变量的方法有三种: 反变换法、组合法、取舍法。其中组合法适合于无法得到反分布函数, 且可以将分布函数分解的情况; 取舍法用于概率密度函数复杂又难以求得其累计分布函数的情况。

a) 反变换法 适合于可以得到反分布函数的模型, 具有直观、快速的特点。因此广泛采用反变换法进行随机抽样。大量试验结果表明, 常用的概率分布有正态分布、对数正态分布和威布尔分布。它们变换公式为:

标准正态分布的反变换公式

$$\begin{aligned} Z_1 &= \sqrt{-2 \ln X_1} \cos(2\pi X_2) \\ Z_2 &= \sqrt{-2 \ln X_1} \sin(2\pi X_2) \end{aligned} \quad (45.3-16)$$

其中  $X_1, X_2$  是  $(0, 1)$  区间上两个独立的均匀分布的随机变量。计算得到的  $Z_1, Z_2$  也是相互独立的随机变量。为了得到服从  $(\mu, \sigma^2)$  分布的随机变量, 需进行如下变换

$$Z_i \times \sigma + \mu, \quad (i = 1, 2) \quad (45.3-17)$$

威布尔分布的反变换公式

$$Z = b(-\ln X)^{\frac{1}{k}} + a \quad (45.3-18)$$

式中  $a$ ——位置参数;

$b$ ——尺度参数;

$k$ ——形状参数;

$X$ —— $(0, 1)$  区间均匀分布的随机变量。

对数正态分布的反变换公式

$$Z = \exp(\sigma t + \mu) \quad (45.3-19)$$

式中  $t$ ——为标准正态分布的随机变量;

$\sigma$ ——为对数标准差;

$\mu$ ——对数均值。

b) 取舍法 对于任何一个分布函数  $g(y)$ ,  $y \in (0, 1)$ , 如果想选出一个  $Y$ , 可先求出  $g_0, g_0$  是  $g(y)$  在  $y(0, 1)$  中的最小上限, 然后产生两个平均分布于  $(0, 1)$  间的随机数  $\xi_1$  和  $\xi_2$ , 若有

$$\xi_2 < \frac{g(\xi_1)}{g_0}$$

则  $Y = \xi_1$

否则, 重新产生两个随机数  $\xi_3$  和  $\xi_4$ , 重复上述试验, 直到满足

$$\xi_{i+1} < \frac{g(\xi_i)}{g_0}$$

则  $Y = \xi_i$

且  $Y$  具有  $g(y)$  的分布。

例 已知  $g(y) = \frac{4}{\pi} \frac{1}{1+y^2}$ ,  $y \in (0, 1)$ , 试依据  $g(y)$  求出  $Y(\xi)$ 。

解 因为  $y \geq 0$ , 所以  $\frac{4}{\pi}$  是  $g(y)$  在  $(0, 1)$  区间内的最大值。对于两个连续的  $\xi$ , 不妨设为  $\xi_i$  和  $\xi_{i+1}$ , 若

$$\xi_{i+1} < \frac{\frac{4}{\pi} \frac{1}{1+\xi_i^2}}{\frac{4}{\pi}}$$

即  $\xi_{i+1}(1+\xi_i^2) \leq 1$ , 则  $Y = \xi_i$

否则, 重新产生两个连续的随机数  $\xi_{i+2}$  和  $\xi_{i+3}$ , 重复上面的试验, 求得的  $Y$  具有  $\frac{4}{\pi} \frac{1}{1+y^2}$  的分布形式。

c) 组合法 若  $g(y)$  为一概率分布, 而  $g(y)$  可以分解为一些分布函数之和, 写为

$$g(y) = \int f_n(y) h(z) dz$$

或  $g(y) = \sum f_n(y) P_n$

先根据  $h(z)$  或  $p(n)$  找出  $z$  或  $n$ , 然后再根据  $f_n$  或  $f_n$  去求出  $y$ 。

例 应用组合法对正态分布进行抽样。

解 引入分布

$$f_1(y) = y \exp(-y^2/2) \quad 0 \leq y$$

$$f_2(y) = \frac{1}{\pi} \sqrt{1-x^2}, \quad -1 \leq x \leq 1$$

则正态分布可以表示为如下的组合分布形式

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_2\left(\frac{x}{y}\right) f_1(y) dy$$

于是有如下正态分布的组合抽样法

$$X_f = \sqrt{-2 \ln \xi_1} \cos 2\pi \xi_2$$

例 已知一零件的应力分布和强度分布都为正态分布, 其数据为:  $\bar{x}_1 = 94.1 \text{ MPa}$ ,  $s_1 = 20.7 \text{ MPa}$ ;  $\bar{x}_2 =$

188.2MPa,  $s_2 = 15.2\text{MPa}$ 。试分别用联结方程及蒙特卡罗法计算其可靠度,并加以比较。

解

1)两个正态分布函数,用下面联结方程求可靠度

$$z = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{s_2^2 + s_1^2}} = \frac{188.2 - 94.1}{\sqrt{20.7^2 + 15.2^2}} = 3.66$$

由表 45.2-4 查得可靠度  $R = 0.99987$ 。

2)由图 45.3-9 的计算机框图编制计算机程序,并得出下列打印结果:

模拟次数	可靠度
1000	0.9990
5000	0.9990
10000	0.9998
50000	0.99978

可见两种计算方法所得的结果比较接近。随着模拟试验次数的增加,模拟结果精度也随之提高。

### 2.4.5 有多种失效模式的可靠度

前面讨论零件的可靠度时只考虑一种失效模式,因此只有一种工作能力的强度判据,但是在实际中常常存在多种失效模式的零件。例如,对于传动螺旋轴,轴承受的载荷过高,使螺杆截面发生断裂、塑性变形,同时还会丧失稳定性而发生纵向弯曲三种失效模式;螺纹牙的失效可能有弯曲、剪切和磨损三种失效模式,因此讨论多种失效模式可靠度具有现实的意义。

假设零件有两种失效模式,第一种失效的失效概率为  $F_1$ ,第二种失效的失效概率为  $F_2$ ,两种同时失效的失效概率为  $F_{12}$ ,则该零件的可靠度的一般表达式为

$$R = 1 - F_1 - F_2 + F_{12} \quad (45.3-20)$$

如果两种失效模式是独立的,则可靠度为

$$R = 1 - F_1 - F_2 + F_1 F_2 \quad (45.3-21)$$

若零件具有三种失效模式,设第  $i$  种失效的失效概率为  $F_i (i=1,2,3)$ ;第  $i$  种失效与第  $j$  种失效同时发生的失效概率为  $F_{ij} (i,j=1,2,3;i \neq j)$ ;第一、二、三种同时失效的概率为  $F_{123}$ ,则可靠度为

$$R = 1 - F_1 - F_2 - F_3 + F_{12} + F_{23} + F_{13} - F_{123} \quad (45.3-22)$$

## 3 可靠度的置信度和置信区间

### 3.1 置信度

置信区间估计,是估计母体的待估计参量(例如母体均值  $\mu$ )以某一给定概率落在多大的区间之内。设母体待估计的参数为  $\theta$ ,给定的概率为  $1-\gamma$ ,区间的下限、上限分别为  $L(x), U(x)$ ,则根据上述定义可表示为

$$P[L(x) \leq \theta \leq U(x)] = 1 - \gamma$$

式(45.3-22)表示母体待估计的参量  $\theta$  的真值落在区间  $[L(x), U(x)]$  内的概率为  $1-\gamma$ 。区间  $[L(x), U(x)]$  称为置信区间,  $1-\gamma$  称为置信度或置信水平或写成  $C=1-\gamma$ 。

### 3.2 单侧置信区间下限和最低可靠度

可靠度的置信区间有三种(参见图 45.3-10)即:真实的可靠度值落在双侧置信区间  $[R_{L2}, R_{U2}]$  内的概率;真实的可靠度值小于单侧置信区间上限  $R_{U1}$  的概率;和真实的可靠度值大于单侧置信区间下限  $R_{L1}$  的概率。在可靠性设计中,常用的是第三种情况,因为就机械零件的强度和寿命而言,关心的是保证零件的最低可靠度。

图 45.3-10 所示三种情况的置信度  $C=1-\gamma$  的表达式分别为

$$C = P(R_{L2} \leq R \leq R_{U2}) = 1 - \gamma \quad (45.3-23)$$

$$C = P(R \leq R_{U1}) = 1 - \gamma \quad (45.3-24)$$

$$C = P(R \geq R_{L1}) = 1 - \gamma \quad (45.3-25)$$

式中  $C$ ——置信度;

$R_{L2}$ ——可靠度的双侧置信区间的下限;

$R_{U2}$ ——可靠度的双侧置信区间的上限;

$R_{U1}$ ——可靠度的单侧置信区间的上限;

$R_{L1}$ ——可靠度的单侧置信区间的下限。

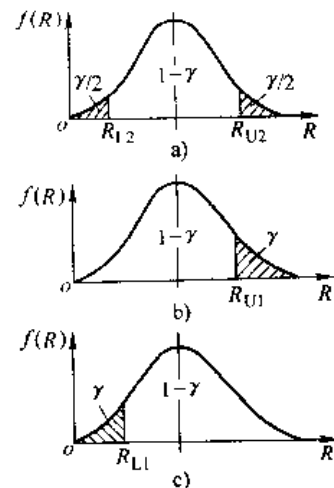


图 45.3-10 可靠度的置信区间

a) 双侧置信区间 b) 单侧置信区间的上限

c) 单侧置信区间的下限

下面只介绍置信区间下限的可靠度。

如前所述,可靠度是  $\delta = (x, -x_1) > 0$  的概率,可写为

$$R = P(\delta > 0) = \int_{-\infty}^{\infty} f(\delta) d\delta$$

如图 45.3-11 所示。



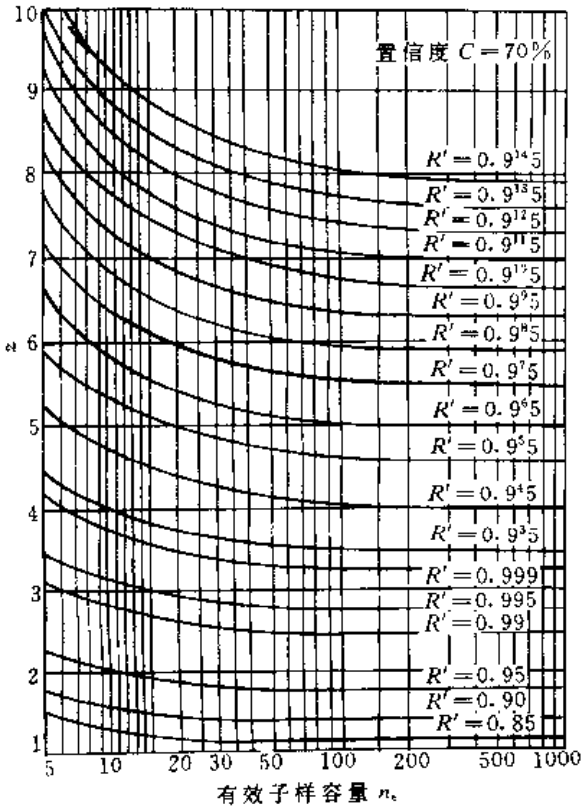


图 45.3-14 70%置信度下的可靠度

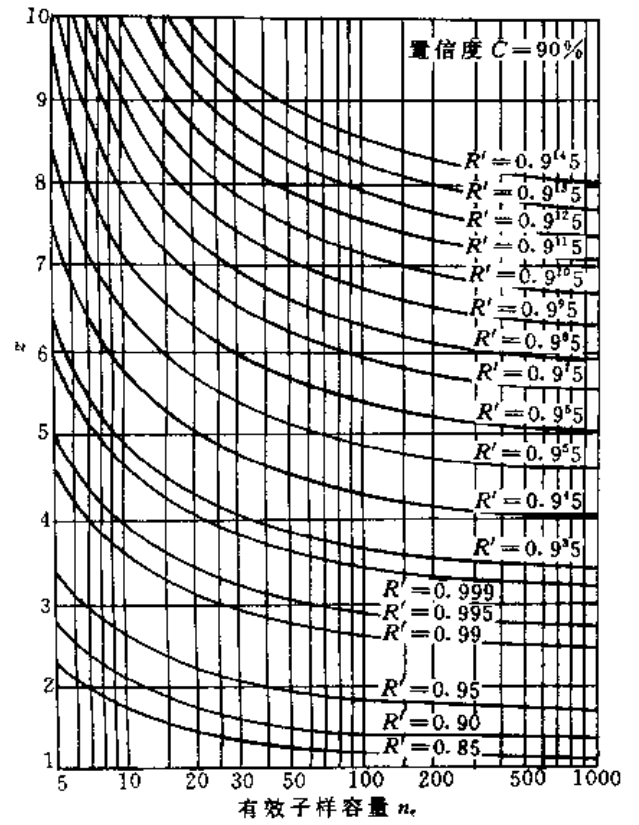


图 45.3-16 90%置信度下的可靠度

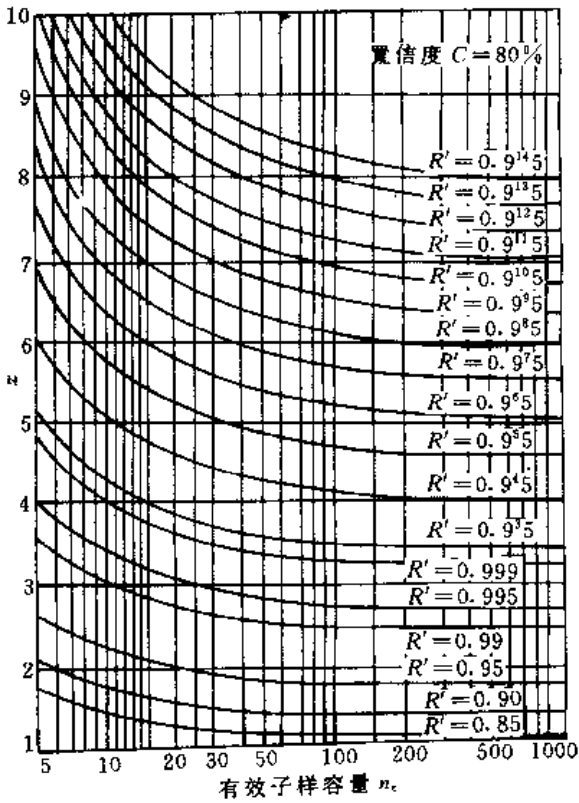


图 45.3-15 80%置信度下的可靠度

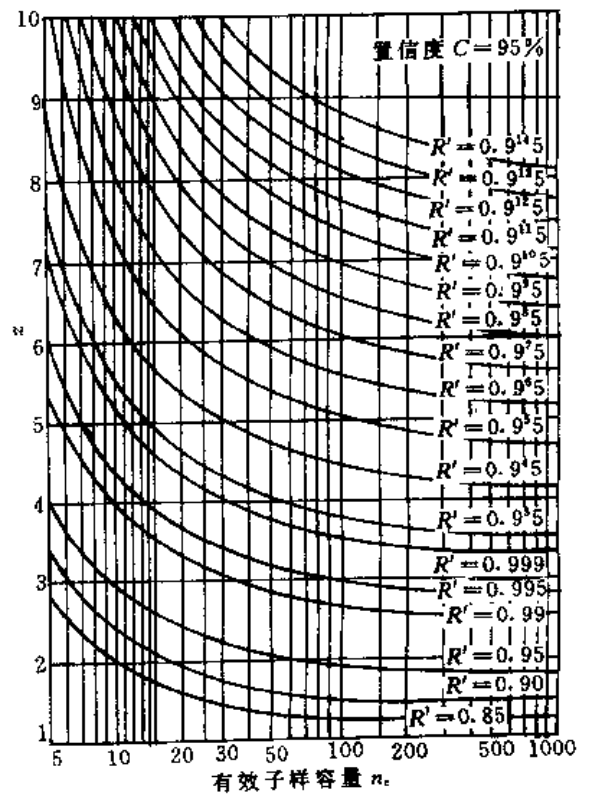


图 45.3-17 95%置信度下的可靠度

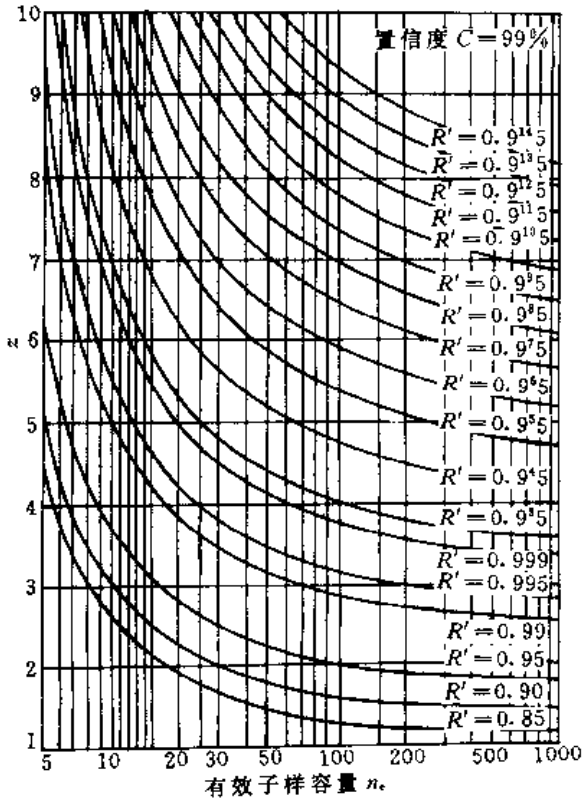


图 45.3-18 99%置信度下的可靠度

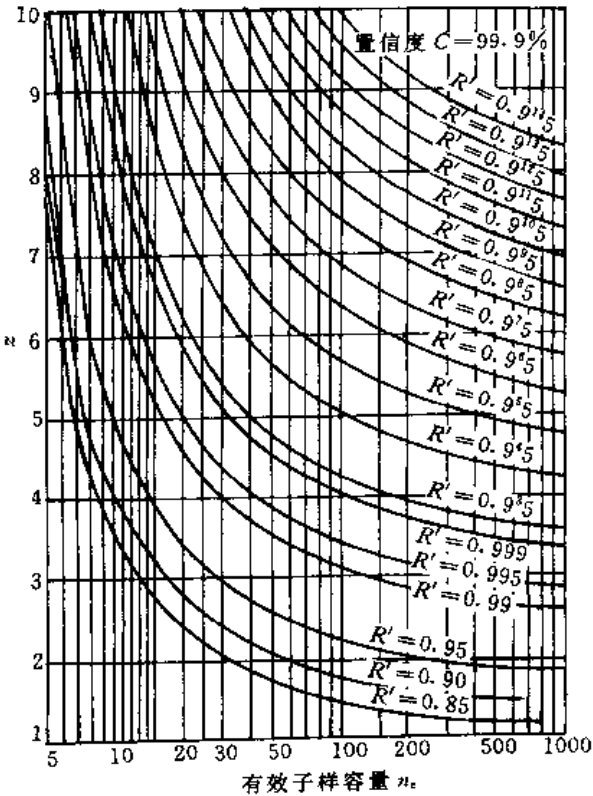


图 45.3-19 99.9%置信度下的可靠度

解 由表(45.3-5)中序号 1 的联结方程

$$z = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{s_1^2 + s_2^2}} = \frac{1213 - 920}{\sqrt{41.2^2 + 81.2^2}} = 3.218$$

由式(45.3-29)

$$n = \frac{(41.2^2/40 + 81.2^2/24)^2}{(41.2^2/40)^2 - (81.2^2/24)^2} - 1 = 31.23$$

由图 45.3-12~45.3-19, 得与所要求的置信度相对应的  $\bar{R}$  及  $R'$  如下

置信度 C (%)	可靠度均值 $\bar{R}$	可靠度的单侧置信下限 $R'$
50	0.99932	0.99932
80	0.99932	0.9965
90	0.99932	0.9945
99	0.99932	0.985

#### 4 随机变量函数的均值和标准差的近似计算

应力和强度常为一些随机变量的函数。已知各随机变量的均值和标准差, 求其函数均值和标准差的近似方法如下。

##### 4.1 泰勒展开法

设  $n$  维随机变量  $x_1, x_2, \dots, x_n$  的函数为

$$y = f(x_1, x_2, \dots, x_n) \quad (45.3-30)$$

各随机变量的变异系数  $v_i = \frac{s_i}{\bar{x}_i}$  很小, 则函数的均值、函数的标准差

各随机变量的变异系数  $v_i = \frac{s_i}{\bar{x}_i}$  很小, 则函数的均值、函数的标准差

$$\begin{aligned} \bar{y} &\approx f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n) + \frac{1}{2} \sum_{i=1}^n \left( \frac{\partial^2 y}{\partial x_i^2} \right)_0 s_{x_i}^2 \\ &+ \sum_{i=1}^{n-1} \times \sum_{j=i+1}^n \left( \frac{\partial^2 y}{\partial x_i \partial x_j} \right)_0 \rho_{ij} s_{x_i} s_{x_j} \\ &\approx f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n) \end{aligned} \quad (45.3-31)$$

$$s_y \approx \left[ \sum_{i=1}^n \left( \frac{\partial y}{\partial x_i} \right)_0^2 s_{x_i}^2 + 2 \sum_{i=1}^{n-1} \times \sum_{j=i+1}^n \left( \frac{\partial y}{\partial x_i} \right)_0 \left( \frac{\partial y}{\partial x_j} \right)_0 \rho_{ij} s_{x_i} s_{x_j} \right]^{\frac{1}{2}} \quad (45.3-32)$$

式中 脚标“0”——求导后自变量取均值;

$\rho_{ij}$ —— $x_i$  与  $x_j$  的相关系数

$$\rho_{ij} = \frac{E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)]}{[E(x_i - \bar{x}_i)^2 E(x_j - \bar{x}_j)^2]^{\frac{1}{2}}} \quad (45.3-33)$$

对  $(x_i, x_j)$  作  $n$  次独立观测的相关系数估计值。一般  $\rho_{ij} = -1 \sim +1$ 。若  $x_i = ax_j + b, a \neq 0$ , 则  $x_i$  与  $x_j$  线性相关。当  $a > 0$ , 则  $\rho_{ij} = +1$ ; 当  $a < 0$ , 则  $\rho_{ij} = -1$ 。若  $x_i$  与  $x_j$  互相独立, 则  $\rho_{ij} = 0$ 。 $\rho_{ij}$  的具体数值, 就概念上判定为  $+1, -1, 0$  常比较容易, 而对非线性又不独立往往

难以估计。在工程计算中,为了简单常假定为独立而取  $\rho_{ij}=0$ , 也有时折中取  $\rho_{ij}=0.5$ 。为避免由此产生的麻烦,有时可设法避开相关的麻烦。

### 4.2 变异系数法

对于单项式(没有加减运算的式子)的函数,式(45.3-30)的具体形式为

$$y = a \prod_{i=1}^n x_i^{m_i} = ax_1^{m_1} x_2^{m_2} \dots x_n^{m_n} \quad (45.3-34)$$

式中  $a, m_i, \dots$  任意常数,可为正也可为负函数的均值。

$$\bar{y} = a \prod_{i=1}^n \bar{x}_i^{m_i} = a \bar{x}_1^{m_1} \bar{x}_2^{m_2} \dots \bar{x}_n^{m_n} \quad (45.3-35)$$

函数的变异系数

$$\nu_y \approx \left( \sum_{i=1}^n m_i^2 \nu_{x_i}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n m_i m_j \rho_{ij} \nu_{x_i} \nu_{x_j} \right) \quad (45.3-36)$$

函数的标准差

$$s_y = \bar{y} \nu_y \quad (45.3-37)$$

### 4.3 基本函数法

本方法是将常用的函数作为基本函数列于表45.3-8,应用时可直接搬用,对于较复杂的函数一般可化为这些基本函数的形式。

表 45.3-8 基本函数形式和近似结论式

序号	函数形式	均值 $\bar{y}$	标准差 $s_y$
1	$y=a$	$a$	0
2	$y=ax$	$a\bar{x}$	$as_x$
3	$y=a \pm x$	$a \pm \bar{x}$	$s_x$
4	$y=r^m$	$\bar{x}^m$	$ m  \bar{x}^{m-1} s_x$
5	$y=r_1 \pm r_2$	$\bar{x}_1 \pm \bar{x}_2$	$(s_{x_1}^2 + s_{x_2}^2 \pm 2\rho_{12}s_{x_1}s_{x_2})^{\frac{1}{2}}$
6	$y=x_1 x_2$	$\bar{x}_1 \bar{x}_2 + \rho_{12}s_{x_1}s_{x_2}$	$(\bar{x}_1^2 s_{x_2}^2 + \bar{x}_2^2 s_{x_1}^2 + 2\rho_{12}\bar{x}_1\bar{x}_2 s_{x_1}s_{x_2})^{\frac{1}{2}}$
7	$y=\frac{x_1}{x_2}$	$\frac{\bar{x}_1}{\bar{x}_2} + \frac{\bar{x}_1 s_{x_2}}{\bar{x}_2^2} \left( \frac{s_{x_2}}{\bar{x}_2} - \frac{\rho_{12}s_{x_1}}{\bar{x}_1} \right) \approx \frac{\bar{x}_1}{\bar{x}_2}$	$\frac{\bar{x}_1}{\bar{x}_2} \left( \frac{s_{x_1}^2}{\bar{x}_1^2} + \frac{s_{x_2}^2}{\bar{x}_2^2} - \frac{2\rho_{12}s_{x_1}s_{x_2}}{\bar{x}_1\bar{x}_2} \right)^{\frac{1}{2}}$

三种求函数均值和标准差的近似法中,基本函数法只在简单情况才显得方便;变异系数法计算最简便,建议尽量采用;泰勒展开法可用于各式各样的函数形式。

例 某钢制拉杆,截面直径  $(\bar{d}, s_d) = (10, 0.1)$  mm, 杆长  $(\bar{L}, s_L) = (1000, 3.3)$  mm, 受拉力  $(\bar{F}, s_F) = (40000, 2000)$  N, 弹性模数  $(\bar{E}, s_E) = (20600, 618)$  MPa, 求拉杆受力后的伸长量。

解 由材料力学知,伸长量

$$\delta = \frac{4FL}{\pi d^2 E}$$

1)用泰勒展开法 用式(45.3-31)式(45.3-32)各随机变量相互独立,  $\rho_{ij}=0$ ,

$$\bar{\delta} = \frac{4\bar{F}\bar{L}}{\pi\bar{d}^2\bar{E}} = \frac{4 \times 40000 \times 1000}{\pi \times 10^2 \times 20600} = 24.72\text{mm}$$

$$s_\delta = \left[ \left( \frac{\partial \delta}{\partial F} \right)_0^2 s_F^2 + \left( \frac{\partial \delta}{\partial L} \right)_0^2 s_L^2 + \left( \frac{\partial \delta}{\partial d} \right)_0^2 s_d^2 + \left( \frac{\partial \delta}{\partial E} \right)_0^2 s_E^2 \right]^{\frac{1}{2}}$$

$$= \left[ \left( \frac{4\bar{L}}{\pi\bar{d}^2\bar{E}} \right)^2 s_F^2 + \left( \frac{4\bar{F}}{\pi\bar{d}^2\bar{E}} \right)^2 s_L^2 + \left( \frac{-8\bar{F}\bar{L}}{\pi\bar{d}^3\bar{E}} \right)^2 s_d^2 + \left( \frac{-4\bar{F}\bar{L}}{\pi\bar{d}^2\bar{E}^2} \right)^2 s_E^2 \right]^{\frac{1}{2}}$$

$$= \left[ \left( \frac{4 \times 1000}{\pi \times 10^2 \times 20600} \right)^2 2000^2 + \left( \frac{4 \times 40000}{\pi \times 10^2 \times 20600} \right)^2 3.3^2 + \left( \frac{-8 \times 40000 \times 1000}{\pi \times 10^3 \times 20600} \right)^2 0.1^2 + \left( \frac{-4 \times 40000 \times 1000}{\pi \times 10^2 \times 20600^2} \right)^2 618^2 \right]^{\frac{1}{2}} \text{mm}$$

$$= 1.526\text{mm}$$

即

$$(\bar{\delta}, s_\delta) = \frac{4(\bar{F}, s_F)(\bar{L}, s_L)}{\pi(\bar{d}, s_d)^2(\bar{E}, s_E)}$$

$$= (24.72, 1.526)\text{mm}$$

2)用变异系数法 用式(45.3-35~37)  $\rho_{ij}=0$ 。并先求出各随机变量的变异系数

$$\nu_d = \frac{s_d}{\bar{d}} = \frac{0.1}{10} = 0.01$$

$$\nu_L = \frac{s_L}{\bar{L}} = \frac{3.3}{1000} = 0.0033$$

$$\nu_F = \frac{s_F}{\bar{F}} = \frac{2000}{40000} = 0.05$$

$$\nu_E = \frac{s_E}{\bar{E}} = \frac{618}{20600} = 0.03$$



则

$$\begin{aligned}\bar{\delta} &= \frac{4}{\pi} \overline{FL} \bar{d}^{-2} \bar{E}^{-1} = \frac{4}{\pi} \times 4000 \times \\ &1000 \times 10^{-2} \times 20600^{-1} \text{mm} \\ &= 24.72 \text{mm}\end{aligned}$$

$$\begin{aligned}s_{\delta} &= (\nu_F^2 + \nu_L^2 + 2^2 \nu_d^2 + \nu_E^2)^{\frac{1}{2}} \\ &= (0.05^2 + 0.0033^2 + 2^2 \times 0.01^2 + \\ &0.03^2)^{\frac{1}{2}} \\ &= 0.06173\end{aligned}$$

$$s_{\delta} = \bar{\delta} \nu_{\delta} = 24.72 \times 0.06173 \text{mm} = 1.526 \text{mm}$$

即  $(\bar{\delta}, s_{\delta}) = (24.72, 1.526) \text{mm}$

3) 用基本函数法 由于变量较多, 故分步计算, 按表 45.3-8

$d^2$ , 按序号 4

$$\bar{d}^2 = \bar{d}^2 = 10^2 = 100$$

$$s_{d^2} = 2\bar{d}\nu_d = 2 \times 10 \times 0.1 = 2$$

$d^2 E$ , 按序号 6,  $\rho_{12} = 0$

$$\overline{d^2 E} = \bar{d}^2 \bar{E} = 10^2 \times 20600 = 2060000$$

$$\begin{aligned}s_{d^2 E} &= [(\bar{d}^2)^2 s_E^2 + \bar{E}^2 s_{d^2}^2]^{\frac{1}{2}} \\ &= [(100)^2 \times 0.03^2 + 20600^2 \times 2^2]^{\frac{1}{2}} \\ &= 74274\end{aligned}$$

$FL$ , 按序号 6,  $\rho_{12} = 0$

$$\overline{FL} = \bar{F} \bar{L} = 40000 \times 1000 = 40000000$$

$$\begin{aligned}s_{FL} &= (\bar{F}^2 s_L^2 + \bar{L}^2 s_F^2)^{\frac{1}{2}} \\ &= (40000^2 \times 3.3^2 + 1000^2 \times 2000^2)^{\frac{1}{2}} \\ &= 2004351\end{aligned}$$

$\frac{FL}{d^2 E}$ , 按序号 7,  $\rho_{12} = 0$

$$\begin{aligned}\frac{\overline{FL}}{\overline{d^2 E}} &= \frac{\overline{FL}}{\overline{d^2 E}} = \frac{40000000}{2060000} \\ &= 19.4175\end{aligned}$$

$$\begin{aligned}s\left(\frac{FL}{d^2 E}\right) &= \frac{\overline{FL}}{\overline{d^2 E}} \left( \frac{s_{FL}^2}{\overline{FL}^2} + \frac{s_{d^2 E}^2}{\overline{d^2 E}^2} \right)^{\frac{1}{2}} \\ &= \frac{40000000}{2060000} \left( \frac{2004351^2}{40000000^2} + \right. \\ &\left. \frac{74274^2}{2060000^2} \right) = 1.1987\end{aligned}$$

$\delta = \frac{4}{\pi} \frac{FL}{d^2 E}$ , 按序号 2

$$\bar{\delta} = \frac{4}{\pi} \left( \frac{FL}{d^2 E} \right) = \frac{4}{\pi} \times 19.4175 \text{mm} = 24.72 \text{mm}$$

$$s_{\delta} = \frac{4}{\pi} s\left(\frac{FL}{d^2 E}\right) = \frac{4}{\pi} \times 1.1987 \text{mm} = 1.526 \text{mm}$$

即  $(\bar{\delta}, s_{\delta}) = (24.72, 1.526) \text{mm}$

## 5 典型机械零件的可靠性设计

### 5.1 螺纹联接的可靠性设计

#### 5.1.1 松螺栓联接

在松螺栓联接中, 螺栓不受预紧力, 只受轴向的随机静载荷。假设拉应力  $\sigma_1$  沿螺栓横截面均匀分布, 失效模式为断裂, 则设计准则为

$$P(\sigma_1 - \sigma_1 > 0) \geq R(t) \quad (45.3-38)$$

对于规定的轴向静载荷、可靠度和材料, 计算螺栓直径及其公差; 确定零件可靠度对断面变化的敏感度。

例 已知松螺栓联接的载荷为:  $(\bar{F}, s_F) = (26700, 900) \text{N}$ , 螺栓材料为 40Cr, 其强度为:  $(\bar{\sigma}_1, s_{\sigma_1}) = (900, 72) \text{MPa}$ 。从经济效益出发, 在保用期内, 10000 个螺栓只允许更换 13 个。设计此松螺栓联接。

解

1) 由载荷引起的应力为

$$\sigma_1 = \frac{F}{A} = \frac{4F}{\pi d^2}$$

由于载荷的统计量已给出, 为求螺栓的应力, 只需确定面积  $A$  的统计量, 由表 45.3-1 得螺栓横截面积  $A$  的均值为

$$\bar{A} = \frac{\pi \bar{d}^2}{4} = f(\bar{d})$$

由式(45.3-32)可得面积  $A$  的标准差为

$$s_A = \left[ \left( \frac{\partial A}{\partial d} \right)^2 s_d^2 \right]^{1/2} = \left[ \left( \frac{\pi \bar{d}}{2} \right)^2 s_d^2 \right]^{1/2} = \frac{\pi \bar{d}}{2} \cdot s_d$$

于是, 预测的应力值为

$$(\bar{\sigma}_1, s_1) = \left( \frac{\bar{F}, s_F}{\bar{A}, s_A} \right) = \left( \frac{26700, 900}{\left( \frac{\pi \bar{d}^2}{4}, \frac{\pi \bar{d}}{2} s_d \right)} \right) \quad (45.3-39)$$

式中, 由于  $\bar{d}$  和  $s_d$  为未知量, 所以  $\bar{\sigma}_1$  和  $s_1$  为未知量。因此需要另一个关系式来确定  $\bar{d}$  和  $s_d$ , 通常最方便的方法是找出制造过程中  $\bar{d}$  与  $s_d$  的关系。由制造公差的统计数据, 可取

$$s_d \approx 0.001 \bar{d} \quad (45.3-40)$$

由式(45.3-39)可得

$$\bar{\sigma}_1 = \frac{4\bar{F}}{\pi \bar{d}^2} = \frac{4 \times 26700}{\pi \bar{d}^2} = \frac{34013}{\bar{d}^2}$$

由表 45.3-1 可得

$$s_1^2 = \frac{\bar{F} s_A^2 + \bar{A}^2 s_F^2}{\bar{A}^4} = \frac{131639817}{\bar{d}^4}$$

$$s_1 = \frac{1147}{\bar{d}^2}$$

于是, 应力的统计量为  $(\bar{\sigma}_1, s_1) = \left( \frac{34013}{\bar{d}^2}, \frac{1147}{\bar{d}^2} \right) \text{MPa}$ ;

2) 静强度的统计量为  $(\bar{\sigma}_s, s_s) = (900, 72) \text{MPa}$ ;

3) 由联结方程确定螺栓尺寸。

松螺栓联接所需的可靠度为

$$R(t) = 1 - \frac{13}{10000} = 0.9987 = \int_x^{\infty} \phi(z) dz$$

由标准正态分布表可知  $z = -3.00$ , 将有关数据代入联结方程, 得

$$-3.00 = \frac{900 - \frac{34013}{\bar{d}^2}}{\left[72^2 + \left(\frac{1147}{\bar{d}^2}\right)^2\right]^{1/2}}$$

化简和整理后, 得

$$\bar{d}^4 - 80.204207\bar{d}^2 + 1499.90 = 0$$

解上式, 得

$$\bar{d} = 5.45 \text{mm}$$

此值应为螺栓的抗拉危险截面上的直径  $d_0$ , 故取螺栓直径为

$$\bar{d} = 8 \text{mm}, \text{内径 } \bar{d}_1 = 6.627 \text{mm}$$

如果按常规设计方法, 则松螺栓联接的强度条件为

$$\frac{4F}{\pi d_c^2} \leq [\sigma] \quad (45.3-41)$$

式中  $[\sigma] = \frac{\sigma_s}{n}$  —— 螺栓的许用拉应力。

安全系数

$$\begin{aligned} n &= \frac{2200k_m}{900 - (70000 - F_0)^2 \times 10^{-7}} \\ &= \frac{2200 \times 1.25}{900 - (70000 - 26700)^2 \times 10^{-7}} \\ &= 4.86 \end{aligned}$$

式中 对于合金钢, 取系数  $k_m = 1.25$ ;

$F_0$  —— 螺栓总拉力, 此处  $F_0 = F = 26700 \text{N}$ 。

对于 40Cr, 屈服点  $\sigma_s = 750 \text{MPa}$

故  $[\sigma] = \frac{750}{3.86} = 194.3 \text{MPa}$ , 代入式 (45.3-41)

$$d_c = \left( \frac{4 \times 26700}{194.3} \right)^{1/2} \text{mm} = 13.22 \text{mm}$$

应取螺栓直径  $d = 16 \text{mm}$ , 内径  $d_1 = 13.835 \text{mm}$ 。

由此可见, 对于松螺栓联接, 按可靠性设计所得的螺栓直径比按常规设计方法所得的直径要小得多。

### 5.1.2 紧螺栓联接

有预紧力和受轴向动载荷的紧螺栓联接, 是螺栓联接中最重要的一种形式。比较典型的是图 45.3-20 所示的发动机气缸盖螺栓联接。

紧螺栓联接的一般设计步骤是:

1) 确定设计准则 假设每个螺栓的拉应力为沿横截面均匀分布, 但由于载荷分布、动态应力集中系数

和几何尺寸等因素的变异性, 对于很多螺栓来说, 每个螺栓的应力值的大小是不一样的, 而是呈分布状态。通常假设它为正态分布。

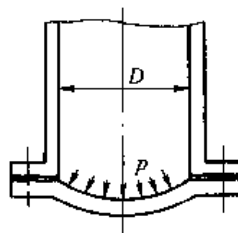


图 45.3-20 受拉紧螺栓联接

对于有紧密性要求的螺栓联接, 假设其失效模式是螺栓发生屈服。因此, 设计准则为: 螺栓材料的屈服点大于螺栓应力  $\sigma_1$  的概率必须大于或等于设计所要求的可靠度  $R(t)$ , 表示为

$$P(\sigma_s > \sigma_1) = P(\sigma_s - \sigma > 0) \geq R(t) \quad (45.3-42)$$

2) 选择螺栓材料 确定其强度分布。根据经验, 可取螺栓拉伸强度的变异系数为

$$\nu_s = 5.3\% \sim 7\%$$

3) 确定螺栓的应力分布。

4) 应用联结方程, 确定螺栓的直径。

例 如图 45.3-20 所示, 已知气缸内直径  $D = 380 \text{mm}$ , 缸内工作压力  $p = 0 \sim 1.70 \text{MPa}$ , 螺栓数目  $n = 8$ , 采用金属垫片, 设计此气缸盖螺栓。要求螺栓联接的可靠度为 0.999999。

解

1) 螺栓材料选用 45 钢, 螺栓性能级别选用 6.8 级, 假设其强度分布为正态分布, 则  $\bar{\sigma}_s = 600 \text{MPa}$ ,  $\bar{\sigma}_s = 480 \text{MPa}$ , (相当于 45 钢正火, 查《材料性能手册》)。

屈服强度的标准差为

$$s_s = 0.07\bar{\sigma}_s = 0.07 \times 480 \text{MPa} = 33.6 \text{MPa}$$

2) 假设螺栓的应力分布为正态分布, 则气缸盖所受最大工作载荷的均值为

$$\bar{F}_T = \bar{p}_{\max} \left( \frac{\pi \bar{d}^2}{4} \right) = 1.70 \times \frac{3.14 \times 380^2}{4} \text{N} = 192700 \text{N}$$

每个螺栓所受最大工作载荷的均值为

$$\bar{F} = \frac{F_T}{n} = \frac{192700}{8} \text{N} = 24090 \text{N}$$

工作载荷的变异系数, 可取  $\nu_F = \frac{s_F}{\bar{F}} = 0.08$ , 因此,

工作载荷分布的标准差

$$s_F = 0.08\bar{F} = 0.08 \times 24090 \text{N} = 1927 \text{N}$$

每个螺栓由工作载荷引起的应力均值为

$$\bar{\sigma}_F = \frac{\bar{F}}{A} = \frac{24090}{\frac{\pi}{4}(\bar{d})^2} = \frac{30688}{\bar{d}^2}$$

应力分布的标准差为

$$s_r = 0.08\bar{\sigma}_r = 0.08 \times \frac{30688}{\bar{d}^2} = \frac{2455}{\bar{d}^2}$$

有预紧力的、受轴向载荷的紧螺栓联接在工作时，螺栓的总拉力为

$$F_0 = F + F' \quad (45.3-43)$$

或 
$$F_0 = \frac{C_1}{C_1 + C_2} F + F' \quad (45.3-44)$$

式中  $F$ ——螺栓上所受的工作载荷；

$F'$ ——预紧力；

$F'_r$ ——剩余预紧力；

$C_1$ ——螺栓刚度系数；

$C_2$ ——被联接件刚度系数；

$\frac{C_1}{C_1 + C_2}$ ——相对刚度。

令  $\frac{C_2}{C_1} = B$ ，则式(45.3-44)可改写为

$$F_0 = \frac{1}{1+B} F + F' \quad (45.3-45)$$

将上式除以螺栓横截面积  $A$ ，可得螺栓总拉应力分布的均值

$$\bar{\sigma} = \frac{F_0}{A} = \frac{1}{1+B} \bar{\sigma}_F + \bar{\sigma}' \quad (45.3-46)$$

当预紧应力  $\sigma_1$  与螺栓的强度成一定比例时，可达到一定的可靠度。根据经验，取预紧应力分布的均值  $\bar{\sigma}_1 = 0.50\sigma_s = 0.5 \times 480\text{MPa} = 240\text{MPa}$ ，标准差  $s_1 = 0.15\bar{\sigma}_1 = 0.15 \times 240\text{MPa} = 36\text{MPa}$ 。

实际上，螺栓的刚度系数  $C_1$  可以较精确地算出 ( $C_1 = \frac{\pi d^2 E}{4l}$ )，而被联接件的刚度系数  $C_2$  则较难精确确定。此处可取比例系数  $B=8$ ，而其变异系数  $v_B = 0.10$ ，所以  $B$  的标准差为  $s_B = 0.10B = 0.10 \times 8 = 0.8$ 。

将有关数值代入式(45.3-46)得

$$\bar{\sigma} = \frac{1}{1+8} \times \frac{30688}{\bar{d}^2} + 240 = \frac{3410}{\bar{d}^2} + 240$$

3) 应用联结方程求螺栓直径 联结方程(表 45.3-5 中序号 1)为

$$z = \frac{\bar{\sigma}_s - \bar{\sigma}}{(s_s^2 + s^2)^{1/2}} = \frac{\mu_s}{s_s} \quad (45.3-47)$$

表 45.3-9 预紧应力  $\sigma_1$  对可靠度  $R(t)$  的影响

比值 $\sigma_1/\sigma_s$	预紧应力 $\sigma_1$ /MPa	螺栓总应力 $\sigma = \frac{C_1}{C_1+C_2}\sigma_r + \sigma_1$ /MPa	螺栓总拉力 $F_0 = \sigma A$ /N	预紧力 $F_1 = \sigma_1 A$ /N	剩余预紧力 $F'_r = F_0 - F$ /N	比值 $\frac{F'_r}{F}$	联结系数 $z = \frac{\sigma_s - \sigma}{s_s}$	可靠度 $R(t)$
0.20	96	104	35157	32454	11067	0.459	-10.28	>0.928
0.40	192	200	67612	54908	13522	1.81	-6.33	0.938
0.50	240	248	89838	61134	9748	2.48	-4.71	0.958
0.60	288	298	100065	97361	95971	3.15	-3.36	0.936
0.70	336	344	116292	113588	92202	3.83	-2.28	0.988
0.80	384	392	132519	129815	125725	4.39	-1.32	0.908

注： $\sigma_s = 480\text{MPa}$ ， $\frac{C_1}{C_1 + C_2} = \frac{1}{1+B} = \frac{1}{9}$ ，工作载荷  $F = 24090\text{N}$ ，螺栓直径  $d = 24\text{mm}$

由式(45.3-47)和式(45.3-46)可知，联结系数  $z$  与四个随机变量有关，它们是  $\sigma_s$ 、 $\sigma_F$ 、 $B$  和  $\sigma_1$ ，对于多维随机变量，由式(45.3-32)，得

$$s_z^2 = \sum_{i=1}^n \left( \frac{\partial z}{\partial x_i} \right)^2 s_{x_i}^2 = \left( \frac{\partial z}{\partial \sigma_s} \right)^2 s_{\sigma_s}^2 + \left( \frac{\partial z}{\partial \sigma_F} \right)^2 s_{\sigma_F}^2 + \left( \frac{\partial z}{\partial B} \right)^2 s_B^2 + \left( \frac{\partial z}{\partial \sigma_1} \right)^2 s_{\sigma_1}^2$$

式中  $\left( \frac{\partial z}{\partial \sigma_s} \right)^2 = 1$ ， $\left( \frac{\partial z}{\partial \sigma_F} \right)^2 = -\frac{1}{1+B}$ ， $\left( \frac{\partial z}{\partial B} \right)^2 = \frac{\sigma_F}{(1+B)^2}$ ， $\left( \frac{\partial z}{\partial \sigma_1} \right)^2 = -1$

代入前式后，得

$$s_z^2 = (1)^2(33.6)^2 + \left( \frac{-1}{1+8} \right)^2 \left( \frac{2455}{\bar{d}^2} \right)^2 + \left[ \frac{1}{(1+8)^2} \right]^2 \left( \frac{30688}{\bar{d}^2} \right)^2 (0.8)^2 + (-1)(36)^2 = \frac{66272}{\bar{d}^4} + 2525$$

$$s_z = \left( \frac{166272}{\bar{d}^4} + 2425 \right)^{1/2}$$

由标准正态分布表 45.2-4 知，当要求可靠度  $R = 0.96$  时， $z = 4.70$ ，于是，将有关各值代入联结方程，得

$$4.70 = \frac{480 - \left( \frac{3410}{\bar{d}^2} + 240 \right)}{\left( \frac{166272}{\bar{d}^4} + 2425 \right)^{1/2}}$$

化简和整理后，得

$$\bar{d}^4 - 405.95\bar{d}^2 + 1973 = 0$$

解上式，得

$$\bar{d} = 20\text{mm}$$

因此，螺栓的尺寸确定如下：

公称直径  $d = 24\text{mm}$ ，内径  $d_1 = 20.752\text{mm}$ ，对上述设计结果讨论如下：

1) 由式(45.3-46)和式(45.3-47)可以看出，当螺栓材料、几何尺寸、工作载荷和联接的相对刚度一定时，螺栓的预紧应力  $\sigma_1$  是影响可靠度大小的主要因素。不同的  $\sigma_1/\sigma_s$  值，对应着不同的可靠度，如表 45.3-9 所示。

由表 45.3-9 可以看出,当螺栓材料选定后,预紧应力  $\sigma_i$  不宜过大,否则会引起可靠度的降低。一般,取  $\sigma_i = (0.40 \sim 0.60)\sigma_s$ 。

2) 表 45.3 10 给出了在改变相对刚度时相应的可靠度。可以看出,在其他条件不变时,即使相对刚度成倍地变化,可靠度  $R(t)$  的变化也不显著。

表 45.3-10 相对刚度  $\frac{C_1}{C_1+C_2} \left( = \frac{1}{1+B} \right)$  对可靠度  $R(t)$  的影响

( $\sigma_s = 480\text{MPa}$ , 工作载荷  $F = 24090\text{N}$ , 螺栓直径  $d = 24\text{mm}$ , 预紧应力  $\sigma_i = 0.50\sigma_s = 240\text{MPa}$ )

相对刚度 $\frac{C_1}{C_1+C_2} = \frac{1}{1+B}$	螺栓总应力 $\sigma = \frac{C_1}{C_1+C_2}\sigma_p + \sigma_i$	螺栓总拉力 $F_n = \sigma A$ /N	剩余预紧力 $F_i' = F_0 - F$ /N	比值 $\frac{F_i'}{F}$	联结系数 $\alpha = \frac{\sigma_s - \sigma}{\sigma_s}$	可靠度 $R(t)$
$\frac{1}{9} = 0.111$	248	83838	59748	2.18	-4.71	0.9-8
$\frac{1}{5} = 0.2$	254	85867	61777	2.56	-4.50	0.9-6
$\frac{1}{2.5} = 0.4$	268	90600	66510	2.76	-4.22	0.9-8
$\frac{1}{1.25} = 0.8$	297	100404	78814	3.17	-3.54	0.9-8

3) 由式 (45.3-46) 和式 (45.3-47) 可知,螺栓材料的强度对于联接的可靠度有一定影响。由表 45.3-9 可知,即使选择强度较低的材料,只要保持一定的  $\sigma_i/\sigma_s$  值,仍可达到较高的可靠度。但是,低强度螺栓不可能达到较大的预紧力  $F$  和剩余预紧力  $F_i'$ 。因此,对于有紧密性要求的螺栓联接,不应选用强度低材料。

4) 这里采用的可靠性设计方法是以螺栓发生屈服作为失效判据。如果以螺栓的疲劳断裂作为设计准则,则由表 45.3-10 可知,不论相对刚度  $\frac{C_1}{C_1+C_2} \left( = \frac{1}{1+B} \right)$  多大;螺栓的应力幅都比较小。由此按联结方程计算出的螺栓直径相当小,即用屈服判据进行可靠性设计是合理的。

### 5.1.3 受剪螺栓联接

例 一受剪螺栓联接如图 45.3-21 所示,已知载荷为等幅循环载荷,呈正态分布,其均值及标准差为  $(\bar{L}, s_L) = (24000, 1440)\text{N}$ , 承切面数  $n=2$ 。预紧力忽略不计。从安全考虑,在 10000 个螺栓中,只允许有两个失效。设计此螺栓联接。

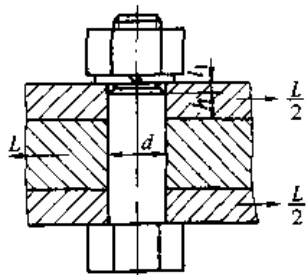


图 45.3-21 受剪螺栓联接

解

(1) 按栓杆受剪切设计

1) 确定失效判据 假设所有的设计变量为正态分

布,失效模式是在循环载荷下螺栓的疲劳,所以设计准则为:螺栓的切应力小于剪切疲劳强度的概率必须大于或等于设计所要求的可靠度  $R(t)$ ,表示为

$$P(\tau_{-1} > \tau) = P(\tau_{-1} - \tau > 0) \geq R(t) \quad (45.3-48)$$

式中  $\tau_{-1}$  —— 螺栓材料的剪切疲劳极限(MPa);

$\tau$  —— 单个螺栓的切应力(MPa)。

2) 选择螺栓材料,确定其强度分布 螺栓材料选用 45 钢,螺栓性能级别选用 6.8 级,假设其强度分布为正态分布,则  $\bar{\sigma}_b = 600\text{MPa}$ ,  $\bar{\sigma}_s = 480\text{MPa}$ 。

按经验公式,得

$$\begin{aligned} \bar{\sigma}_i &= 0.23(\bar{\sigma}_b + \bar{\sigma}_s) \\ &= 0.23(600 + 480)\text{MPa} = 248.4\text{MPa} \\ \bar{\tau}_{-1} &= 0.577\bar{\sigma}_i = 0.577 \times 248.4\text{MPa} = 143.3\text{MPa} \\ s_{\tau_{-1}} &= 0.08\bar{\tau}_{-1} = 0.08 \times 143.3\text{MPa} = 11.5\text{MPa} \end{aligned}$$

3) 确定螺栓的切应力分布 实际上只有无螺纹的栓杆部分承受切变载荷,所以承受切面积是按栓杆部分算出的。

载荷和栓杆横截面面积为独立变量,故切应力为

$$(\bar{\tau}, s_\tau) = \left( \frac{\bar{L}, s_L}{\bar{A}, s_A} \right) \quad (45.3-49)$$

$$\text{式中 } A = \frac{\pi d^2}{4}$$

由随机变量代数(表 45.3-1),当  $z = x^2$ ,  $z$  的标准差为

$$\sigma_z = 2\mu_z \sigma_x$$

故面积  $A$  的标准差为

$$s_A = \frac{\pi d s_d}{2}$$

式(45.3-49)中的  $\bar{\tau}$  和  $s_\tau$  为未知量,因为  $A$  和  $s_A$  为未知量。为了求出  $A$  和  $s_A$ ,必须找出另一个关系式;一般是根据螺栓直径的制造公差统计量,找出  $d$  与  $s_d$

之间的关系式,有

$$s_d = 0.002d$$

将有关各值代入式(45.3-49),得切应力的均值

$$\bar{\tau} = \frac{\bar{L}}{nA} = \frac{4\bar{L}}{n\pi\bar{d}^2} = \frac{4 \times 2400}{2 \times 3.14 \times \bar{d}^2} = \frac{15287}{\bar{d}^2}$$

切应力的标准差为

$$s_\tau = \frac{4}{n\pi} \left[ \frac{\bar{L}^2(0.004\bar{d}^2)^2 + (\bar{d}^2)^2(s_d)^2}{(\bar{d}^2)^4} \right]^{1/2}$$

$$= \frac{4}{2 \times 3.14} \left[ \frac{24000^2 \times (0.004\bar{d}^2)^2 + (\bar{d}^2)^2(1440)^2}{(\bar{d}^2)^4} \right]^{1/2}$$

$$= \frac{919}{\bar{d}^2}$$

式中  $s_d^2 = 2\bar{d}s_d = 2\bar{d} \times 0.002\bar{d} = 0.004\bar{d}^2$

4) 按联结方程求螺栓直径 由标准正态分布表知,当  $R(t) = 0.9998$ ,联结系数  $z = 3.50$ ,将有关各值代入联结方程,得

$$3.50 = \frac{143.3 - \left( \frac{15287}{\bar{d}^2} \right)}{\left[ 11.5^2 + \left( \frac{919}{\bar{d}^2} \right)^2 \right]^{1/2}}$$

化简和整理后,得

$$\bar{d}^4 - 231.53\bar{d}^2 + 11808 = 0$$

解上式,得栓杆直径为

$$\bar{d} = 12.5\text{mm}$$

圆整后,取标准值,得

螺栓公称直径  $d = 16\text{mm}$ ,内径  $d_1 = 13.835\text{mm}$ 。

### 5.1.4 按栓杆或孔壁受挤压进行设计

由于受剪切螺栓除在联接接合面上受剪切之外,还与被联接件的孔壁互相挤压,所以栓杆或孔壁被压溃也是一种主要失效模式。特别是当被联接件的材料强度较低(例如材料为铸铁或铝合金)时,被联接件的孔壁被压溃常常是主要的失效模式。

(1) 按被联接件孔壁被压溃考虑,设计准则为

$$P(\sigma_q < S_q) \geq R(t) \quad (45.3-50)$$

式中  $S_q$ ——被联接件的挤压强度,可取  $S_q = 0.5\sigma_b$

$\sigma_q$ ——被联接件孔壁上的挤压应力。

$\sigma_q$  与杆孔配合、零件变形、表面加工等有关,难以精确地确定,通常可假设其为均匀分布,如图 45.3-22 所示。于是可得

$$\sigma_q = \frac{L}{ndh} \quad (45.3-51)$$

式中  $\frac{L}{n}$ ——被联接件所受的切应力。

(2) 选择被联接件的材料,确定挤压强度分布

被联接件选用灰铸铁,其抗拉强度为  $\sigma_b = 245\text{MPa}$ 。由经验公式,对于铸铁,可取

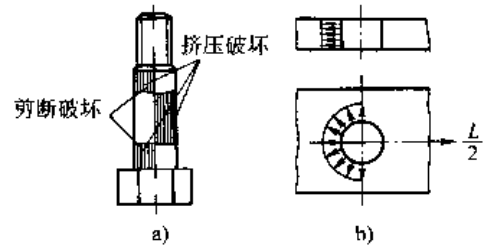


图 45.3-22 被联接件孔壁的假设的挤压应力分布  
a) 螺栓的失效型式 b) 被联接件孔壁的挤压应力

$$\bar{S}_q = 0.5\sigma_b = 0.5 \times 245\text{MPa} = 122.5\text{MPa}$$

$$\text{标准差为 } \sigma_{s,q} = 0.08\bar{S}_q = 0.08 \times 122.5\text{MPa} = 9.8$$

MPa

(3) 确定螺栓的应力分布

为了求  $\bar{\sigma}_q$  和  $s_q$ , 必须求出承受挤压的当量面积  $A = dh$ 。为了方便,可以先找出  $d$  与  $h$  之间的关系。这里按最危险的情况考虑,取  $h = 0.5d$ ,于是可得

$$\bar{\sigma}_q = \frac{2\bar{L}}{n\bar{d}^2} = \frac{2 \times 24000}{2 \times \bar{d}^2} = \frac{24000}{\bar{d}^2}$$

$$s_q \approx \frac{2}{n} \left[ \frac{\bar{L}^2(0.004\bar{d}^2)^2 + (\bar{d}^2)^2 s_L^2}{(\bar{d}^2)^4} \right]^{1/2}$$

$$= \frac{1443}{\bar{d}^2}$$

(4) 应用联结方程求螺栓直径

按题意所要求的可靠度为  $R = 0.9998$ ,与此相应的联结系数  $z = 3.50$ ,故由联结方程可得

$$3.50 = \frac{122.5 - \left( \frac{24000}{\bar{d}^2} \right)}{\left[ 9.8^2 + \left( \frac{1443}{\bar{d}^2} \right)^2 \right]^{1/2}}$$

化简和整理后,得

$$\bar{d}^4 - 425.17\bar{d}^2 + 39805 = 0$$

解上式,得栓杆直径为

$$\bar{d} = 16.9\text{mm}$$

圆整后,取螺栓的公称直径  $d = 20\text{mm}$ ,螺栓内径  $d_1 = 17.294\text{mm}$ 。可见,当尺寸  $h$  较小时( $h = 0.5d$ ),按抗压强度所需的螺栓直径比按栓杆抗剪强度所需的螺栓直径要大。如果取尺寸  $h = 0.75d$  并按上述步骤计算,可得

$$\bar{d} = 13.9\text{mm}$$

圆整后,取螺栓的公称直径  $d = 16\text{mm}$ ,内径  $d_1 = 13.835\text{mm}$ 。可见,这一公称直径与按栓杆抗剪强度求得的公称直径相同。如果取尺寸  $h = d$ ,则得  $\bar{d} = 13.4\text{mm}$ ,圆整后,取螺栓的公称直径  $d = 16\text{mm}$ 。

所以,为使螺栓联接的抗压强度不至于太低,应使尺寸  $h$  不小于  $0.75d$ 。

### 5.2 过盈联接的可靠性设计

图 45.3-23 所示的过盈联接零件内的应力为两向应力状态。其可靠性设计准则为：当量单向应力小于强

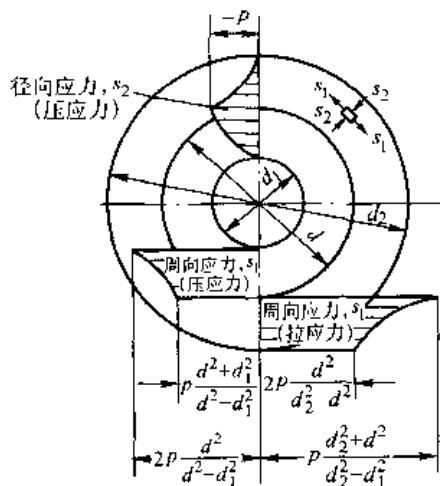


图 45.3-23 过盈联接零件内的应力

度的概率应大于或等于可靠度，表示为

$$P(S > \sigma_e) \geq R(t) \quad (45.3-52)$$

式中  $S$ ——强度，可取  $S = \sigma_e$  (屈服点)；

$\sigma_e$ ——当量单向应力。

按最大变形能理论，对于受两向应力的零件，当平均应力  $\sigma_{1m} = \sigma_{2m} = 0$ ，则

$$\sigma_e = (\sigma_{1a}^2 + \sigma_{2a}^2 - \sigma_{1a}\sigma_{2a})^{1/2} \quad (45.3-53)$$

由表 45.3-1，得

$$\bar{\sigma}_e = (\bar{\sigma}_{1a}^2 + \bar{\sigma}_{2a}^2 - \bar{\sigma}_{1a}\bar{\sigma}_{2a})^{1/2}$$

式中  $\sigma_a$ ——应力幅。

若  $\sigma_{1a}$  和  $\sigma_{2a}$  是相关的，相关系数  $\rho = 1$ ，则有

$$\frac{\sigma_{2a}}{\sigma_{1a}} = c$$

代入式(45.3-53)，得

$$\sigma_e \approx \sigma_{1a}(c^2 - c + 1)^{1/2}$$

由表 45.3-1 得当量应力的统计量为

$$\left. \begin{aligned} \bar{\sigma}_e &\approx \bar{\sigma}_{1a}(c^2 - c + 1)^{1/2} \\ \bar{s}_e &\approx \bar{s}_{1a}(c^2 - c + 1)^{1/2} \end{aligned} \right\} \quad (45.3-54)$$

对于包容件，危险应力发生在内表面处，该处的周向应力

$$\sigma_{1a} = p \frac{d_2^2 + d^2}{d_2^2 - d^2}$$

径向应力

$$\sigma_{2a} = -p$$

对于被包容件，危险应力也发生在内表面处，该处的周向应力

$$\sigma_{1a} = \frac{2pd^2}{d^2 - d_1^2}$$

径向应力

$$\sigma_{2a} = 0$$

例 一蜗轮以过盈联接装在轴上，并用平键联接作为辅助，如图 45.3-24 所示。联接传递的转矩  $T = 1000\text{N} \cdot \text{m}$ ，轴向力  $F = 2500\text{N}$ 。轴和毂的尺寸如图所示，蜗轮轮芯的材料为铸钢 ZG310—570，其屈服点  $\sigma_{s2} = 320\text{MPa}$ ；轴的材料为 45 钢，其屈服点  $\sigma_{s1} = 360\text{MPa}$ 。轴和孔的表面粗糙度的  $R_a$  值为  $0.80\mu\text{m}$ ，拟用压入法装配。设已知最大径向压力  $p_{\text{max}} = 117.8\text{MPa}$ ，试对此过盈联接进行可靠性分析。

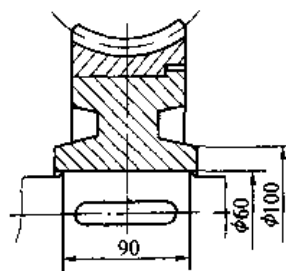


图 45.3-24 蜗轮与轴的过盈联接

解 (1) 对于包容件

周向应力

$$\begin{aligned} \bar{\sigma}_1 &= p \left( \frac{d_2^2 + d^2}{d_2^2 - d^2} \right) = p \left( \frac{100^2 + 60^2}{100^2 - 60^2} \right) \\ &= 2.125\text{MPa} \approx -117.8\text{MPa} \end{aligned}$$

取  $s_1 \approx 0.09\bar{\sigma}_1 = 0.09 \times 250\text{MPa} = 22.5\text{MPa}$

径向应力  $\bar{\sigma}_2 = -p = -117.8\text{MPa}$

取  $s_2 = 0.09\bar{\sigma}_2 = 0.09 \times 250\text{MPa} = 10.6\text{MPa}$

系数  $c = \frac{\bar{\sigma}_1}{\bar{\sigma}_2} = \frac{p}{2.125p} = 0.471$

$$c^2 = 0.222$$

$$(c^2 - c + 1)^{1/2} = (0.222 - 0.471 + 1)^{1/2} = 0.8666$$

故当量单向应力为

$$\begin{aligned} \bar{\sigma}_e &= (c^2 - c + 1)^{1/2} \bar{\sigma}_1 = 0.8666 \times 250\text{MPa} \\ &\approx 217\text{MPa} \end{aligned}$$

$$\begin{aligned} s_e &= (c^2 - c + 1)^{1/2} s_1 = 0.8666 \times 22.5\text{MPa} \\ &\approx 19.5\text{MPa} \end{aligned}$$

包容件的强度取为屈服点，其均值及标准差为

$$\bar{\sigma}_s = 320\text{MPa}$$

取  $s_s = 0.08\bar{\sigma}_s = 0.08 \times 320\text{MPa} = 25.6\text{MPa}$ 。

由联结方程，得联结系数

$$z = \frac{320 - 217}{(25.6^2 + 19.5^2)^{1/2}} = 3.20$$

由标准正态分布表，得可靠度为

$$R(t) = 0.9993$$

(2) 对于被包容件

周向应力  $\bar{\sigma}_1 \approx \frac{2pd^2}{d^2-d_1^2}$ , 令  $d_1=0$ , 故  $\bar{\sigma}_1=0.09 \times$

235.6MPa=21.2MPa; 径向应力  $\bar{\sigma}_2=0$ ,

系数  $c = \frac{\bar{\sigma}_2}{\bar{\sigma}_1} = 0$

故  $\bar{\sigma}_e = \bar{\sigma}_1 = 235.6\text{MPa}$   
 $s_e = s_1 = 21.2\text{MPa}$

被包容件的强度取为屈服点, 其均值及标准差为  
 $\bar{\sigma}_s = 360\text{MPa}$

$s_s = 0.08\bar{\sigma}_s = 0.08 \times 360\text{MPa} = 28.8\text{MPa}$

由联接方程, 得

$$z = \frac{360 - 235.6}{(28.8^2 + 21.2^2)^{1/2}} = -3.48$$

由标准正态分布表, 得可靠度为

$$R(t) = 0.9997$$

### 5.3 压缩螺旋弹簧的可靠性设计

这种弹簧设计的基本问题是: ①满足强度要求, 使切应力不引起失效; ②满足刚度要求, 使弹簧的变形不超过规定量; ③确定所需的有效圈数。螺旋弹簧的主要失效模式是疲劳断裂。相应的设计步骤如下:

(1) 确定失效应力分布

螺旋弹簧中的最大切应力发生在簧丝的内侧, 为

$$\tau = k_s \frac{8FD}{\pi d^3} \quad (45.3-55)$$

式中  $k_s$ ——曲度系数;

$$k_s = 1 + \frac{0.5}{c} = 1 + \frac{d}{2D} \quad (45.3-56)$$

$c$ ——弹簧指数;

$$c = \frac{D}{d} \quad (45.3-57)$$

$F$ ——轴向压力;

$D$ ——弹簧的平均直径;

$d$ ——簧丝直径。

弹簧的变形量为

$$y = \frac{8FD^3n}{d^4G} \quad (45.3-58)$$

力与变形量的关系为

$$F = \frac{d^4G}{8D^3n}y = ky \quad (45.3-59)$$

式中  $k$ ——弹簧比;

$$k = \frac{d^4G}{8D^3n}; \quad (45.3-60)$$

$G$ ——弹簧材料的切变模量;

$n$ ——有效圈数;

$$n = \frac{Gd^4y}{8FD^3} \quad (45.3-61)$$

将式(45.3-59)代入式(45.3-55), 得最大切应力

与变形量的关系为

$$\tau = \frac{\left(1 + \frac{d}{2D}\right)dGy}{\pi D^2n} \quad (45.3-62)$$

由式(45.3-62), 切应力的均值和标准差分别为

$$\bar{\tau} = \frac{\left(1 + \frac{d}{2D}\right)dG\bar{y}}{\pi D^2n}$$

$$s_\tau = \left[ \left(\frac{\partial \bar{\tau}}{\partial d}\right)^2 s_d^2 + \left(\frac{\partial \bar{\tau}}{\partial D}\right)^2 s_D^2 + \left(\frac{\partial \bar{\tau}}{\partial G}\right)^2 s_G^2 + \left(\frac{\partial \bar{\tau}}{\partial y}\right)^2 s_y^2 + \left(\frac{\partial \bar{\tau}}{\partial n}\right)^2 s_n^2 \right]^{1/2}$$

式中  $s_d$ ——簧丝直径的标准差; 对于冷拔碳钢, 可取

$s_d = 0.00158d$ ; 对于不锈钢, 变异系数可取得稍大些;

$s_D$ ——弹簧平均直径的标准差;

对于冷拔碳钢和不锈钢, 可取  $s_D = 0.0058D$

$s_G$ ——切变模量的标准差; 可取  $s_G = 0.02\bar{G}$

$s_y$ ——弹簧变形量的标准差; 可取  $s_y = 0.02\bar{y}$

$s_n$ ——簧圈数的标准差; 总圈数中有 1/4~1/2

圈的变异性, 所以可取  $3s_n = \frac{1}{4}$ , 即:  $s_n = \frac{1}{12}$

$\approx 0.0833$ 。

(2) 确定强度分布

弹簧的强度分布可按经验方法确定, 如图 45.3-25 所示。

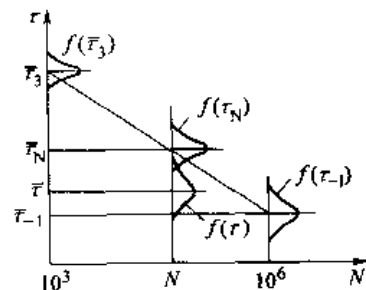


图 45.3-25 剪切疲劳线图

当  $N=10^3$  时的疲劳强度分布为

$$\tau_3 \approx 0.9\tau_b \quad (a)$$

式中  $\tau_b$ ——抗切强度。

$$\tau_b = \beta\sigma_b \quad (b)$$

式中 系数  $\beta \sim N(0.975, 0.05)$ 。

将式(b)和式(c)代入式(a), 得其均值  $\bar{\tau}_3$  及标准差  $s_{\tau_3}$  为

$$\bar{\tau}_3 = 0.9 \times 0.75 \times \bar{\sigma}_b = 0.675\bar{\sigma}_b \quad (d)$$

$$s_{\tau_3} = (\bar{\sigma}_b^2 \times s_\beta^2 + \bar{\beta}^2 \times s_b^2)^{1/2} = (0.0025\bar{\sigma}_b^2 + 0.5625s_b^2)^{1/2} \quad (e)$$

式中  $s_b = 0.05\bar{\sigma}_b$ 。

变异系数

$$\nu_3 = \frac{s_{\tau_3}}{\bar{\tau}_3}$$

当  $N=10^6$  时, 弹簧材料的强度疲劳分布为

$$\tau_0 \approx 0.577\sigma_{-1} - 0.577k_c\sigma'_{-1}$$

式中  $\sigma'_{-1}$ ——试样的疲劳极限, 可取  $\sigma'_{-1} = 0.5\sigma_b$ , 代入上式, 得

$$\bar{\tau}_0 = 0.2885k_c\sigma_b \quad (f)$$

考虑到  $\sigma_0$  的分布, 可取  $s_{\tau_0} =$

$$0.2885k_c s_\sigma$$

$k_c$ ——应力集中修正系数, 由图 45.3-26 给出,  $k_c$  是弹簧常数  $c$  的函数。

变异系数

$$\nu_{\tau_0} = \frac{s_{\tau_0}}{\tau_0}$$

假设当  $10^3 \leq N \leq 10^6$  次时剪切疲劳强度  $\tau_N$  的变异系数,  $\nu_N = s_N/\tau_N$  为直线变化, 则

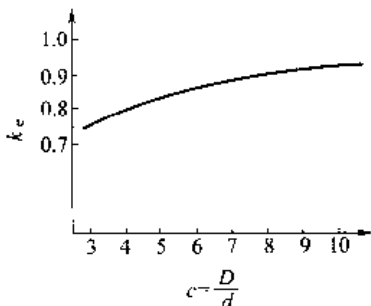


图 45.3-26 弹簧应力集中修正系数  $k_c$

$$\nu_N = \nu_{\tau_3} = \nu_{\tau_6} + \left( \frac{\lg N}{3} - 1 \right) (\nu_{\tau_6} - \nu_{\tau_3}) \quad (45.3-63)$$

同理, 假设在任意寿命  $N$  时的剪切疲劳强度  $\bar{\tau}_N$  按直线变化(对数坐标), 则

$$\begin{aligned} \lg \bar{\tau}_N &= \lg \bar{\tau}_3 + \left( \frac{\lg N}{3} - 1 \right) (\lg \bar{\tau}_6 - \lg \bar{\tau}_3) \\ &= \lg \left[ 0.675\sigma_b + \left( \frac{\lg N}{3} - 1 \right) \right] [\lg \bar{\tau}_6 \\ &\quad - \lg \bar{\tau}_3 (0.0675\sigma_b)] \end{aligned} \quad (45.3-64)$$

当  $N \geq 10^6$  时,  $\tau_N = \tau_6 = \tau_{-1}$

例 一不锈钢螺旋压缩弹簧的参数如表 45.3-11 试确定下列问题:

表 45.3-11 螺旋压缩弹簧的参数

参数	均值	标准差
$d$	2.0mm	0.008mm
$D$	16.0mm	0.0928mm
$G$	79250MPa	1585MPa
$n$	14	0.0833
$y$	20mm	0.4mm
$\sigma_b$	1667MPa	84MPa

1) 在设计寿命为  $N=5 \times 10^4$  循环时的可靠度;

2) 在设计寿命为  $N \geq 10^6$  循环时的可靠度;

3) 弹簧比  $k$  的分布。

解 (1) 确定  $N=5 \times 10^4$  循环时的可靠度

1) 切应力的均值和标准差分别为

$$\begin{aligned} \bar{\tau} &= \frac{\left( 1 + \frac{\bar{d}}{2D} \right) \bar{d} G \bar{y}}{\pi \bar{D}^2 \bar{n}} \\ &= \frac{\left\{ 1 + \frac{2.0}{2 \times 16.0} \right\} \times 2/0 \times 60250 \times 20.0}{3.14 \times 16.0^2 \times 14} \text{MPa} \\ &= 299.3 \text{MPa} \end{aligned}$$

$$\begin{aligned} s_\tau &= \left[ \left( \frac{\partial \bar{\tau}}{\partial \bar{d}} \right)^2 s_d^2 + \left( \frac{\partial \bar{\tau}}{\partial D} \right)^2 s_D^2 + \left( \frac{\partial \bar{\tau}}{\partial G} \right)^2 s_G^2 \right. \\ &\quad \left. + \left( \frac{\partial \bar{\tau}}{\partial y} \right)^2 s_y^2 + \left( \frac{\partial \bar{\tau}}{\partial n} \right)^2 s_n^2 \right] \end{aligned}$$

$$\text{式中 } \frac{\partial \bar{\tau}}{\partial \bar{d}} = \frac{G \bar{y}}{\pi \bar{n}} \left( \frac{1}{D^2} + \frac{\bar{d}}{D^3} \right)$$

$$= \frac{79250 \times 20}{3.14 \times 14} \left( \frac{1}{16^2} + \frac{2}{16^3} \right) = 158.5$$

$$\frac{\partial \bar{\tau}}{\partial D} = \frac{G \bar{y}}{\pi \bar{n}} \left[ (-2) \frac{\bar{d}}{D^3} + (-3) \frac{\bar{d}^2}{2D^4} \right]$$

$$= \frac{79250 \times 20}{3.14 \times 14} = \left[ (-2) \frac{2}{16^3} + (-3) \frac{2^2}{2 \times 16^4} \right] = -38.5$$

$$\frac{\partial \bar{\tau}}{\partial G} = \frac{\bar{y}}{\pi \bar{n}} \left[ \frac{\bar{d}}{D^2} + \frac{\bar{d}^2}{2D^3} \right] = \frac{79250 \times 20}{3.14 \times 14} \left( \frac{2}{16^2} + \frac{2^2}{2 \times 16^3} \right) = 0.00377$$

$$\frac{\partial \bar{\tau}}{\partial y} = \frac{\bar{y}}{\pi \bar{n}} \left[ \frac{\bar{d}}{D^2} + \frac{\bar{d}^2}{2D^3} \right] = \frac{79250 \times 20}{3.14 \times 14} \left( \frac{2}{16^2} + \frac{2^2}{2 \times 16^3} \right) = 15.0$$

$$\frac{\partial \bar{\tau}}{\partial n} = (-1) \frac{G \bar{y}}{\pi \bar{n}} \left( \frac{\bar{d}}{D^2} + \frac{\bar{d}^2}{2D^3} \right)$$

$$= (-1) \frac{79250 \times 20}{3.14 \times 14^2} \left( \frac{2}{16^2} + \frac{2^2}{2 \times 16^3} \right) = -21.4$$

关于切应力公式中的五个随机变量, 其标准差分别取为代入上式, 得

$$s_d \approx 0.004\bar{d} = 0.004 \times 2\text{mm} = 0.008\text{mm}$$

$$s_D \approx 0.0058\bar{D} = 0.0058 \times 16\text{mm} = 0.00928\text{mm}$$

$$s_G \approx 0.002\bar{G} = 0.02 \times 79250\text{MPa} = 1585\text{MPa}$$

$$s_y = 0.0058\bar{y} = 0.02 \times 20\text{mm} = 0.4\text{mm}$$

$$s_n = 0.0833$$

代入上式

$$\begin{aligned} s_\tau &= [158.5^2 \times 0.08^2 + (-38.5)^2 \times 0.00928^2 + \\ &\quad 0.00377^2 \times 1585^2 + 15^2 \times 0.4^2 + (-21.4)^2 \times \\ &\quad 0.0833^2]^{1/2} \text{MPa} \\ &= 9.54 \text{MPa} \end{aligned}$$

2) 确定强度分布

不锈钢的强度极限为



$$(\bar{\sigma}_b, s_b) = (1667, 84) \text{MPa}$$

当  $N = 10^5$  时, 由式(d)和式(e)可得

$$\bar{r}_1 = 0.675\bar{\sigma}_b = 0.675 \times 1667 \text{MPa} = 1125 \text{MPa}$$

$$\begin{aligned} s_{r_1} &= (0.0025\bar{\sigma}_b^2 + 0.5625s_b^2)^{1/2} \\ &= (0.0025 \times 1667^2 + 0.5625 \times 84^2)^{1/2} \text{MPa} \\ &= 104.5 \text{MPa} \end{aligned}$$

于是可得

$$\nu_{r_1} = \frac{104.5}{1125} = 0.0929$$

弹簧指数为

$$c = \frac{D}{d} = \frac{16}{2} = 8$$

当  $N = 10^5$  时, 由图 45.3-26,  $k_c = 0.9$ , 由式(f)和式(g)于是可得

$$\begin{aligned} \bar{r}_2 &= 0.2885k_c\bar{\sigma}_b = 0.2885 \times 0.9 \times 1667 \text{MPa} \\ &= 132.8 \text{MPa} \end{aligned}$$

$$\begin{aligned} s_{r_2} &= 0.2885\sigma_k k_c \sigma_b = 0.2885 \times 0.9 \times 84 \text{MPa} \\ &= 21.8 \text{MPa} \end{aligned}$$

于是可得

$$\nu_{r_2} = \frac{21.8}{132.8} = 0.0504$$

由式(45.3-63), 当  $N = 5 \times 10^4$  时切变疲劳强度的变异系数为

$$\begin{aligned} \nu_c &= \nu_{r_1} + \left[ \frac{\lg N}{3} - 1 \right] (\nu_{r_2} - \nu_{r_1}) \\ &= 0.0929 + \left[ \frac{\lg(5 \times 10^4)}{3} - 1 \right] (0.0504 - 0.0929) \\ &= 0.0688 \end{aligned}$$

由式(45.3-64)当  $N = 5 \times 10^4$  时的剪切疲劳强度均值为

$$\begin{aligned} \lg \bar{r}_N &= \lg(0.675 \times \sigma_b) + \left[ \frac{\lg N}{3} - 1 \right] [\lg \bar{r}_2 - \\ &\quad \lg(0.675\bar{\sigma}_b)] \\ &= \lg(0.675 \times 1667) + \left[ \frac{\lg 50000}{3} - 1 \right] [\lg 132.8 - \\ &\quad - \lg(0.675 \times 1667)] \\ &= 2.8162424 \end{aligned}$$

得

$$\begin{aligned} \bar{r}_N &= 655.00 \text{MPa} \\ s_{r_N} &= \bar{r}_N \times \nu_c = 655 \times 0.0688 \text{MPa} = 45.1 \text{MPa} \end{aligned}$$

3) 计算可靠度

将应力分布参数和强度分布参数代入联接方程。

得

$$z = \frac{655 - 299.3}{(45.1^2 + 9.54^2)^{1/2}} = -7.716$$

故  $R(N = 5 \times 10^4) = 0.913$ 。

(2) 确定  $N \geq 10^6$  时的可靠度

当  $N \geq 10^6$  时,  $r_N = r_{-1} = 432.8 \text{MPa}$

由联接方程, 得

$$z = \frac{132.8 - 299.3}{(21.8^2 + 9.54^2)^{1/2}} = -5.610$$

故  $R(N \geq 10^6) = 0.978$

(3) 确定弹簧比  $k$  的分布

设  $k$  为正态分布, 由式(45.3-60), 得

$$\bar{k} = \frac{d^4 G}{8D^3 n} = \frac{2^4 \times 79250}{8 \times 16^3 \times 14} \text{N/mm} = 2.7640 \text{N/mm}$$

$$s_k = \left[ \left( \frac{\partial k}{\partial d} \right)^2 s_d^2 + \left( \frac{\partial k}{\partial D} \right)^2 s_D^2 + \left( \frac{\partial k}{\partial G} \right)^2 s_G^2 + \left( \frac{\partial k}{\partial n} \right)^2 s_n^2 \right]^{1/2}$$

$$\text{式中 } \frac{\partial k}{\partial d} = \frac{1}{2} \frac{d^4 G}{D^3 n} = \frac{1}{2} \frac{2^4 \times 79250}{16^3 \times 14} = 5.52804$$

$$\frac{\partial k}{\partial D} = \frac{1}{24} \frac{d^4 G}{D^4 n} = \frac{1}{24} \frac{2^4 \times 79250}{16^3 \times 14} = 1.71429$$

$$\frac{\partial k}{\partial G} = \frac{d^4}{8D^3 n} = \frac{2^4}{8 \times 16^3 \times 14} = 3.48772 \times 10^{-5}$$

$$\frac{\partial k}{\partial n} = \frac{d^4 G}{8G^2 n^2} = -\frac{2^4 \times 79250}{8 \times 16^3 \times 14} = -2.76402$$

代入上式, 得

$$\begin{aligned} s_k &= [5.52804^2 \times 0.008^2 + (-1.71429)^2 \times \\ &\quad 0.0928^2 + (3.48772 \times 10^{-5})^2 \times 1585^2 + \\ &\quad (-2.76402)^2 \times 0.0833^2]^{1/2} = 0.283 \end{aligned}$$

故弹簧比  $k$  的分布为

$$k \sim N(2.764, 0.283) \text{N/mm}$$

由于  $F = ky$ , 所以在确定了  $k$  的分布之后, 不难求得变形量  $y$  或轴向压力  $F$  的分布。

## 5.4 滚动轴承的可靠性设计

滚动轴承是最早具有可靠性指标的机械零件。现行的滚动轴承计算方法规定, 在等于额定动载荷  $C$  的当量动载荷作用下, 滚动轴承可以工作一百万转而其中 90% 不发生疲劳点蚀失效, 即其可靠度为 90%。

如果要求的可靠度为 0.90, 则可以按额定动载荷的计算方法去计算  $C$ , 并据以选择轴承。如果要求的可靠度不为 0.90, 则应当计算出与目标可靠度  $R(t)$  相应的可靠寿命或额定动载荷, 并据此选择可靠度为 0.90 的轴承。

### 5.4.1 滚动轴承寿命与可靠度之间的关系

滚动轴承的疲劳寿命服从威布尔分布, 轴承在寿命  $t$  时的失效概率为

$$F(t) = 1 - e^{-\left(\frac{t}{k}\right)^b} \quad (45.3-65)$$

式中  $t$ ——轴承寿命;

$b$ ——尺度参数;

$k$ ——形状参数。

因  $F(t) = 1 - R(t)$ , 故可得与  $t$  对应的可靠度为

$$R(t) = e^{-\left(\frac{t}{L}\right)^k} \quad (45.3-66)$$

上式可改写为

$$t = b[-\ln R(t)]^{\frac{1}{k}} \quad (a)$$

当  $R(t) = 0.90$  时, 轴承的寿命  $t = L_{10}$ ,  $L_{10}$  表示失效概率为 10% 的寿命。于是由上式可得

$$L_{10} = b[-\ln 0.9]^{\frac{1}{k}} \quad (b)$$

由式(a)除以式(b), 整理后可得

$$t = L_{10} \left[ \frac{\ln R(t)}{\ln 0.9} \right]^{\frac{1}{k}} \quad (45.3-67)$$

式中  $t$  —— 与  $R(t)$  相应的可靠寿命。

上式的适用范围在  $0.4 < R(t) < 0.93$  之间。按轴承类型的不同, 形状参数  $k$  的值如下:

球轴承	$k = \frac{10}{9}$
滚子轴承	$k = \frac{3}{2}$
圆锥滚子轴承	$k = \frac{4}{3}$

为了考虑不同可靠度对轴承寿命的影响和便于计算, 将式(45.3-67)简化为

$$t = aL_{10} \quad (c)$$

式中  $a$  —— 滚动轴承寿命可靠性系数列于表 45.3-12。

$$a = \left[ \frac{\ln R(t)}{\ln 0.9} \right]^{\frac{1}{k}} \quad (45.3-68)$$

实际上遇到的问题, 常常是给定目标可靠度下的可靠寿命, 需要确定其所对应的额定寿命  $L_{10}$  值, 然后据以在目录中选用轴承。

由式(45.3-68)可得

$$L_{10} = \frac{t}{a}$$

表 45.3-12 滚动轴承寿命可靠性系数  $a$  值

可靠度 $R(t)(\%)$	50	80	85	90	92	95	96	97	98	99
轴承寿命 $L$	$L_{50}$	$L_{20}$	$L_{15}$	$L_{10}$	$L_8$	$L_5$	$L_4$	$L_3$	$L_2$	$L_1$
球轴承	5.45	1.96	1.48	1.00	0.81	} 0.62 0.53 0.44 0.33 0.21				
圆柱滚子轴承	3.51	1.65	1.34	1.00	0.86					
圆锥滚子轴承	4.11	1.76	1.38	1.00	0.84					

例 一 6209 深沟球轴承在某项应用中得出具有 90% 可靠度的疲劳寿命为  $100 \times 10^6$  转。问如果具有 95% 的可靠度时, 疲劳寿命有多大?

解 由式(45.3-67)可得

$$t = L_{10} \left[ \frac{\ln 0.95}{\ln 0.9} \right]^{\frac{1}{k}} = 100 \times 10^6 \times 0.523 = 52.3 \times 10^6 \text{ 转}$$

例 用一对圆柱滚子轴承支承的轴, 要求在系统可靠度为 0.98 时有 1000h 的可靠寿命, 如已知轴的可靠度为  $R_1(t) = 0.999$ , 求在选择这对轴承时应取的额定寿命值。

解 轴与一对轴承属于串联系统, 系统的可靠度为

$$R_s(t) = R_1(t)[R_2(t)]^2$$

每个轴承的可靠度应为

$$R_2(t) = \left[ \frac{R_s(t)}{R_1(t)} \right]^{1/2} = \left( \frac{0.98}{0.999} \right)^{1/2} = 0.990$$

由表 45.3-12 得,  $a = 0.21$ , 故应取的额定寿命为

$$L_{10} = \frac{t}{a} = \frac{1000}{0.21} \text{ h} = 4762 \text{ h}$$

可见, 选择一只可靠度为 90%、寿命为 4762h 的轴承, 如果用于要求可靠度为 0.99 的场合, 其当量寿命仅为 1000h。

### 5.4.2 滚动轴承额定动载荷与可靠度之间的关系

轴承额定动载荷与其寿命之间的关系为

$$L_{10} = \left( \frac{C}{P} \right)^{\epsilon} \quad (45.3-69)$$

式中  $C$  —— 额定动载荷 (N);

$P$  —— 当量动载荷 (N);

$\epsilon$  —— 疲劳寿命系数, 对于球轴承,  $\epsilon = 3$ ; 对于圆柱滚子和圆锥滚子轴承,  $\epsilon = \frac{10}{3}$ 。

考虑到不同的可靠度, 不同的材料和润滑条件; 式(45.3-69)可表示为

$$t = a \cdot b \cdot c \left( \frac{C}{P} \right)^{\epsilon} \quad (45.3-70)$$

式中  $a$  —— 寿命可靠性系数, 列于表 45.3-12;

$b$  —— 材料系数, 对于普通轴承钢,  $b = 1$ ;

$c$  —— 润滑系数, 一般条件下,  $c = 1$ 。

假设取  $b = c = 1$ , 则式(45.3-70)可写为

$$C = a^{-\frac{1}{\epsilon}} P \cdot t^{\frac{1}{\epsilon}} = K P t^{\frac{1}{\epsilon}} \quad (45.3-71)$$

式中  $K$  —— 额定动载荷可靠性系数, 列于表 45.3-13。

$$K = a^{-\frac{1}{\epsilon}} = \left[ \frac{\ln 0.9}{\ln R(t)} \right]^{\frac{1}{k}}$$

对于球轴承,  $\frac{1}{k\epsilon} = \frac{3}{10}$ ;

对于圆柱滚子轴承,  $\frac{1}{k\epsilon} = \frac{1}{5}$ ;

对于圆锥滚子轴承,  $\frac{1}{k\epsilon} = \frac{9}{40}$ 。

表 45.3-13 滚动轴承额定动载荷可靠性系数  $K$  值

可靠度 $R(t)(\%)$	50	80	85	90	92	95	96	97	98	99
轴承寿命 $L$	$L_{50}$	$L_{20}$	$L_{15}$	$L_{10}$	$L_k$	$L_5$	$L_4$	$L_3$	$L_2$	$L_1$
球轴承	0.5683	0.7984	0.8781	1.00	1.073	1.155 1.209 1.282 1.391 1.600				
圆柱滚子轴承	0.6861	0.8606	0.9170	1.00	1.048					
圆锥滚子轴承	0.6545	0.8446	0.9071	1.00	1.054					

当已知目标可靠度下的轴承寿命,即可由式(45.3-71)确定相应的额定动载荷  $C$  值,然后据以选择轴承。

例 一滚子轴承,  $d = 40\text{mm}$ , 受径向力  $F_1 = 6000\text{N}$  作用,要求可靠度  $R(t) = 0.95$ , 工作寿命  $t = 5000\text{h}$ , 选择此轴承。

解 由式(45.3-71)得

$$C = KPt^{\frac{1}{5}} = 1.155 \times 6000 \times 5000^{\frac{1}{5}}\text{N} = 38065\text{N}$$

故可以选择 2308。

如果只要求可靠度  $R(t) = 0.90$ , 工作寿命  $t = L_0 = 5000\text{h}$ , 则由式(45.3-69)得

$$C = PL_{10}^{\frac{1}{5}} = 6000 \times 5000^{\frac{1}{5}}\text{N} = 32957\text{N}$$

故只需选择 2208 即可。

如果只有 2208 可供选用,而又要求可靠度为 0.95,则可以允许的径向力  $F_1$ , 便需降低,由

$$C = KPt^{\frac{1}{5}} = 1.155 \times P \times 5000^{\frac{1}{5}}\text{N} = 35792\text{N}$$

$$\text{可得 } P = \frac{35792}{1.155 \times 5000^{\frac{1}{5}}}\text{N} = 5642\text{N}$$

## 第 4 章 机械系统的可靠性

系统是由某些彼此相互协调工作的零部件、子系统组成的为了完成某一特定功能的综合体。组成系统并相对独立的机件,通称为单元。系统按其可否修复分为不可修复系统和可修复系统两类。系统的可靠性不仅与组成系统的各单元的可靠性有关,还取决于单元的组合方式和单元失效之间的相互关系。

系统可靠性设计的目的是使系统在满足规定的可靠性指标的前提下,协调、优化系统的技术性能、重量指标、制造成本等;或者在性能、重量、成本、寿命和其他要求的约束下,设计出高可靠性系统。

系统可靠性设计包括两部分内容。按照已知零部件或各单元的可靠性数据,计算系统的可靠性指标,称为可靠性预测;按照给定的系统可靠性指标,对组成系统的单元进行可靠性分配。

### 1 不可修复系统的可靠性模型

系统可靠性模型表示系统可靠性与单元可靠性的关系。它由可靠性逻辑框图(简称可靠性框图)和可靠性数学模型组成。可靠性框图是单元正常或失效状态导致系统正常或失效的逻辑关系示意图。可靠性数学模型是可靠性框图的代数描述,它给出系统可靠性与单元可靠性定量的函数关系。

可靠性框图与系统的结构示意图并不相同。例如图 45.4-1 所示为由两个阀门 1 和 2 组成的简单系统。当系统的功能是使流体能通过,则阀门能打开是该单元正常,否则该单元失效。显然,只要阀门 1 和 2 中任一个失效,则流体都不能通过,即系统失效,这时可靠性框图如图 45.4-2a 所示。当系统的功能是能阻止流体通过,则阀门能关闭是该单元正常,否则该单元失效。显然,只有阀门 1 和 2 都失效,流体的通过才不能阻止,即系统失效。这时可靠性框图如图 45.4-2b 所示。下面给出几种典型的系统可靠性模型。

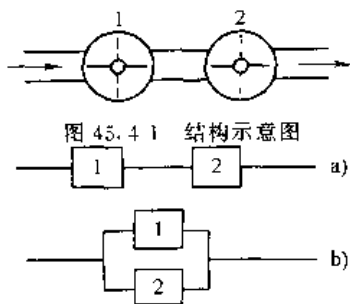


图 45.4-2 可靠性框图

### 1.1 串联系统

设由  $n$  个部件组成的系统,其中任一部件发生故障,系统即出现故障,或者说只有全部部件都正常系统才正常,这样的系统称为串联系统,其可靠性框图如图 45.4-3 所示。

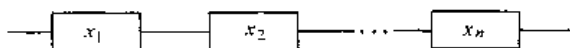


图 45.4-3 串联系统可靠性框图

描述串联系统的概率表达式为

$$R_s = P(A_1 \cap A_2 \cap \dots \cap A_n) \quad (45.4-1)$$

式中  $R_s$ ——系统的可靠度;

$A_i$ ——第  $i$  个零件正常的事件,  $i=1, 2, \dots, n$ ;

$n$ ——系统中零件的总数;

$P(\cdot)$ ——概率。

如果各事件的正常或失效是相互独立的事件,则式(45.4-1)为

$$R_s = P(A_1)P(A_2)\dots P(A_n) = \prod_{i=1}^n P(A_i) \quad (45.4-2)$$

当第  $i$  各单元的故障率函数为  $\lambda$ ,则系统的可靠度为

$$\begin{aligned} R_s(t) &= \prod_{i=1}^n e^{-\int_0^t \lambda_i(u) du} = e^{-\int_0^t \sum_{i=1}^n \lambda_i(u) du} \\ &= e^{-\int_0^t \lambda(u) du} \end{aligned} \quad (45.4-3)$$

由此可见,一个由独立单元组成的串联系统的故障率是所有单元故障率之和,单元数越多,系统故障率越高,可靠度也就越低。

### 1.2 并联系统

设系统  $n$  个部件组成,若至少一个部件正常系统即正常,或必须所有  $n$  个部件都发生故障时系统才出现故障,这样的系统称为并联系统。并联系统的可靠性框图如图 45.4-4 所示。并联系统的概率表达式为

$$\begin{aligned} R_s &= P(A_1 \cup A_2 \cup \dots \cup A_n) \\ &= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) \end{aligned} \quad (45.4-4)$$

假设各单元相互独立,则由概率乘法定理可得系统的可靠度为

$$R_s = 1 - \prod_{i=1}^n [1 - P(A_i)] \quad (45.4-5)$$

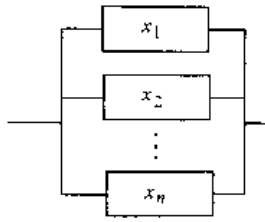


图 45.4-4 并联系统  
可靠性框图

式中  $R_s$  —— 系统可靠度；

$P(i)$  —— 第  $i$  单元可靠度。

上式表明并联系统的可靠度高于任何一个单元的可靠度。

### 1.3 混联系统

#### (1) 串-并联系统

图 45.4-5 是并联子系统的串联方案简称串-并联系统。计算该系统的可靠度时，首先将并联子系统化为一个等效部件，然后将整个系统当作一个串联系统来计算。

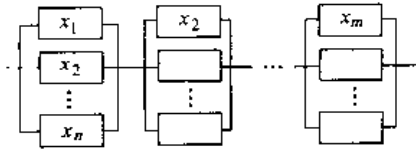


图 45.4-5 串-并联系统

设有  $m$  个子系统，每个子系统都是由  $n$  个部件并联组成的，各个部件的可靠度分别为  $R_{ij}, i=1, 2, \dots, m; j=1, 2, \dots, n$ ，且所有部件的故障都相互独立，则可得串-并联系统的可靠度为

$$R_s = \prod_{i=1}^m \left\{ 1 - \prod_{j=1}^n (1 - R_{ij}) \right\} \quad (45.4-6)$$

当所有的  $R_{ij} = R$ ，所有的  $n_i = n$ ，则  $m$  个并联子系统的串联系统的可靠度为

$$R_s = [1 - (1 - R)^n]^m \quad (45.4-7)$$

#### (2) 并-串联系统的可靠度

并-串系统如图 45.4-6 所示。计算这种系统可靠度的方法是首先将每一串子系统化成一个等效部件，然后把整个系统看作是并联系统来计算。

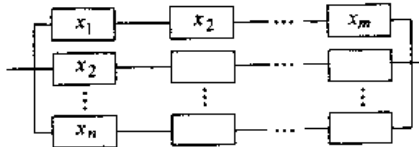


图 45.4-6 并-串系统

假设有  $n$  个子系统，每一子系统有  $m$  个部件，各个部件的可靠度分别为  $R_{ij}, i=1, 2, \dots, m; j=1, 2, \dots, n$ ，且所有部件的故障都相互独立，则可得并-串系

统的可靠度为

$$R_s = 1 - \prod_{i=1}^n \left\{ 1 - \prod_{j=1}^m R_{ij} \right\} \quad (45.4-8)$$

当所有的  $R_{ij} = R$ ，所有的  $n_i = n$ ，则系统的可靠度为

$$R_s = 1 - (1 - R^m)^n \quad (45.4-9)$$

例 如图 45.4-7a、b 所示两系统，若单元可靠度均为  $R=0.75, m=n=4$ ，分别求其系统可靠度。

解 串并联系统，按式(45.4-7)

$$R_s = [1 - (1 - R)^n]^m = [1 - (1 - 0.75)^4]^4 = 0.98447$$

并串联系统，按式(45.4-9)

$$R_s = 1 - (1 - R^m)^n = 1 - (1 - 0.75^4)^4 = 0.78163$$

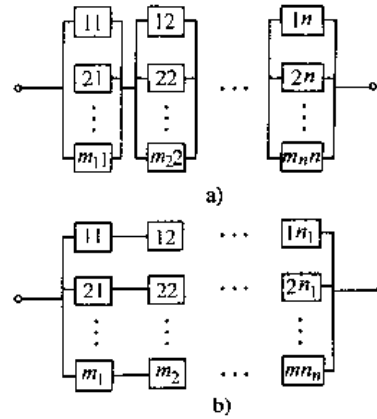


图 45.4-7 串-并联系统  
与并-串系统

### 1.4 表决系统

设系统由  $n$  个部件组成，而系统成功地完成任务需要其中至少  $k$  个部件是好的，这种系统称为  $k/n(G)$  结构，或称  $n$  中取  $k$  表决系统 ( $1 \leq k \leq n$ )，其中  $G$  表示系统完好。例如电力系统、多个发动机的飞机、由多根钢索拧成的钢缆等等，都可以称为  $k/n(G)$  系统。图 45.4-8 所示为典型的  $k/n(G)$  系统原理图。

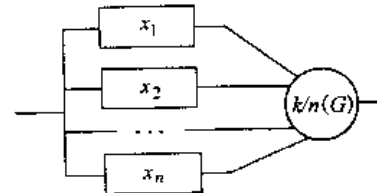


图 45.4-8  $k/n(G)$  原理框图

表决系统有以下三种情况：

1) 当  $k=n, n/n(G)$  系统等价于  $n$  个部件的串联系统；

2) 当  $k=1, 1/n(G)$  系统等价于  $n$  个部件的并联

系统;

3) 当  $k=m+1, m+1/(2m+1), (G)$  系统称为多数表决系统。

$$\text{定义: } x_i = \begin{cases} 1 & \text{第 } i \text{ 个部件正常} \\ 0 & \text{第 } i \text{ 个部件故障} \end{cases}$$

若系统完好必须满足下式

$$\sum_{i=1}^n x_i \geq k \quad (45.4-10)$$

对于  $n$  个部件相同的情况,可借助二项定理求其可靠度。当系统由  $n$  个相同的部件组成时,设某一部件在规定的时间内可靠度为  $R$ ,那么在规定的时间内不可靠度为  $Q$ ,且  $R+Q=1$ ,对于  $n$  个部件来说

$$(R+Q)^n = 1 \quad (45.4-11)$$

将上式展开

$$(R+Q)^n = R^n + C_n^1 R^{n-1} Q + C_n^2 R^{n-2} Q^2 + \dots + Q^n = 1 \quad (45.4-12)$$

即可得到  $k/n(G)$  系统的可靠度为

$$R_s = \sum_{i=k}^n \binom{n}{i} R^i (1-R)^{n-i} \quad (45.4-13)$$

例 如图 45.4-9 所示表决系统,由 6 个相同单元组成,若单元寿命均为指数分布,失效率均为  $40 \times 10^{-6}/h$ ,有不少于 3 个单元工作,系统就不失效,求工作到 7200h 的系统可靠度。

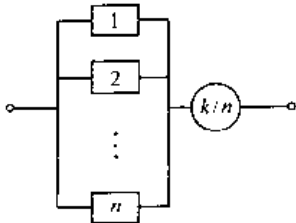


图 45.4-9 表决系统

解 首先求单元可靠度,指数分布

$$R = R(t) = e^{-\lambda t} = e^{-40 \times 10^{-6} \times 7200} = 0.75$$

再用式(45.4-13)求系统可靠度

$$\begin{aligned} R_s &= \sum_{i=3}^6 \binom{6}{i} R^i (1-R)^{6-i} \\ &= \sum_{i=3}^6 \binom{6}{i} 0.75^i (1-0.75)^{6-i} \\ &= \frac{6!}{3!(6-3)!} 0.75^3 (1-0.75)^{6-3} + \\ &\quad \frac{6!}{4!(6-4)!} 0.75^4 (1-0.75)^{6-4} \\ &= \frac{6!}{5!(6-5)!} 0.75^5 (1-0.75)^{6-5} + \\ &\quad \frac{6!}{6!(6-6)!} 0.75^6 (1-0.75)^{6-6} \\ &= 0.9624 \end{aligned}$$

### 1.5 贮备系统的可靠度

贮备系统由  $n$  个部件组成,在初始时刻,一个部件开始工作,其余  $n-1$  个部件作为贮备。当工作部件故障时,贮备部件逐个地替换故障部件,直至所有  $n$  个部件均发生故障,系统才发生故障。贮备系统包括冷贮备系统和热贮备系统,图 45.4-10 为贮备系统的可靠性框图。

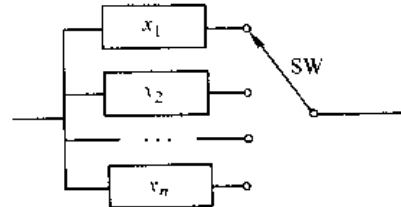


图 45.4-10 贮备系统可靠性框图

贮备系统有冷贮备系统和热贮备系统两种。所谓冷贮备系统是指贮备期间贮备部件不通电,不运行,所以贮备部件故障不劣化,贮备期长短对以后的工作寿命没有影响。但是,贮备系统中贮备部件的替换故障部件的转换开关对整个系统能否可靠正常地工作影响很大,因此根据转换开关 SW 的可靠性,冷贮备系统可以分为两种情况:转换开关完全可靠(理想转换开关)的冷贮备系统和转换开关不完全可靠的冷贮备系统。

因为在热贮备系统中贮备部件在贮备期间可能通电和运转,因此可能发生故障,所以热贮备系统比冷贮备系统复杂的多,贮备寿命与工作寿命分布一般不相同,在这里不作详述。

冷贮备系统中,依开关 SW 的性质又分为理想转换开关和非理想转换开关两种。

(1) 理想转换开关的冷贮备系统

若系统是由  $n$  个指数型部件组成,系统的可靠度为

$$R_s(t) = \sum_{i=1}^n \left[ \prod_{j=1, j \neq i}^n \frac{\lambda_j}{\lambda_j + \lambda_i} \right] \cdot e^{-\lambda_i t} \quad (45.4-14)$$

系统的平均寿命为

$$MTTF_s = \sum_{i=1}^n \frac{1}{\lambda_i} \quad (45.4-15)$$

(2) 转换开关不完全可靠的冷贮备系统

设系统有两个部件组成,转换开关的可靠度为  $R_K(t)$ ,则系统的可靠度  $R_s(t)$  为

$$R_s(t) = \int_0^{\infty} f_1(\tau) d\tau + R_K(t) \int_0^t f_1(\tau_1) \int_{t-\tau_1}^{\infty} f_2(\tau_2) d\tau_2 d\tau_1 \quad (45.4-16)$$

式中  $f_1(\tau)$  —— 部件 1 的概率密度函数;

$f_2(\tau)$  —— 部件 2 的概率密度函数。

如果转换开关和两个部件皆服从指数分布,其失效率分别为  $\lambda_K, \lambda_1, \lambda_2$ ,则可靠度为

$$R_s(t) = e^{-\lambda_1 t} + \frac{\lambda_2}{\lambda_1 + \lambda_2} [e^{-\lambda_1 t} - e^{-(\lambda_1 + \lambda_2)t}] \quad (45.4-17)$$

系统的平均寿命  $MTTF_s = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \cdot \frac{1}{\lambda_1 + \lambda_2}$  (45.4-18)

**例** 已知两单元组成的理想转换开关的冷贮备系统,各单元的失效率  $\lambda_1 = \lambda_2 = 0.0001h^{-1}$ ,求系统工作到 1000h 的可靠度。

**解** 根据式(45.4-14)该系统的可靠度计算式为

$$\begin{aligned} R_s(t) &= e^{-\lambda t} \cdot \sum_{i=0}^n \frac{(\lambda t)^i}{i!} \\ &= e^{-0.0001 \times 1000} \cdot (1 + 0.0001 \times 1000) \\ &= 0.9953 \end{aligned}$$

### 1.6 复杂系统

在实际问题中,有许多系统是不能简单地分解成串联、并联分系统再进行计算。这种系统称为复杂系统,例如,图 45.4-11 是电桥电路系统,称为桥式复杂系统。

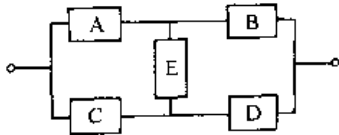


图 45.4-11 桥式复杂系统

复杂系统求可靠度有多种方法,现仅介绍两种。

#### (1) 全概率法

全概率法也称分解法,其数学模型运用全概率公式可得

$$R_s = P(S) = P(X)P(S|X) + P(\bar{X})P(S|\bar{X}) \quad (45.4-19)$$

式中  $S$ ——事件,系统正常;

$X$ ——事件,系统中单元  $X$  正常;

$\bar{X}$ ——事件,系统中单元  $X$  失效;

$P(S|X)$ ——单元  $X$  正常(犹如短路)时系统正常的条件概率;

$P(S|\bar{X})$ ——单元  $X$  失效(犹如断路)时系统正常的条件概率。

运用式(45.4-19)的关键是适当的选择单元  $X$ 。例如图 45.4-11 的桥式复杂系统,若选择  $E$  单元正常为式(45.4-19)中事件  $X$ ,显然,

$$P(E) = R_E$$

$$P(\bar{E}) = 1 - R_E$$

$P(S|E)$ ,当  $E$  处短路,相当于两对并联  $A、C$  和  $B、D$  的串并联,即

$$P(S|E) = [1 - (1 - R_A)(1 - R_C)] \times [1 - (1 - R_B)(1 - R_D)]$$

$P(S|\bar{E})$ ,当  $E$  处断路,相当于两对串联  $A、B$  和  $C、D$  的并串联,即

$$P(S|\bar{E}) = 1 - [1 - (R_A R_B)(1 - R_C R_D)]$$

将这些关系代入式(45.4-19)即得该系统

$$\begin{aligned} R_s &= R_E [1 - (1 - R_A)(1 - R_C)] \times \\ &\quad [1 - (1 - R_B)(1 - R_D) + (1 - R_E)] \times \\ &\quad \{1 - [(1 - R_A R_B)(1 - R_C R_D)]\} \end{aligned}$$

#### (2) 穷举列表法

穷举列表法也称布尔真值表法。这是最直观的方法。各单元只有失效和不失效两种状态,因此  $n$  个单元共有  $2^n$  种状态,将这  $2^n$  种状态全部列出制成穷举表(也称布尔真值表),如表 45.4-1 所示。该表有  $2^n$  行,  $n$  列,表中每一栏填入数“0”(代表单元失效)或“1”(代表单元不失效)。系统用“ $F$ ”代表失效,用“ $S$ ”代表不失效。将系统所有不失效状态的概率相加,即为系统的可靠度。显然,当单元数  $n$  较多,则系统组合状态数  $2^n$  将相当大,用本法求可靠度就很繁琐。

**例** 如图 45.4-12 所示系统,若  $R_A = 0.9, R_B = 0.8, R_C = 0.7, R_D = 0.6$ ,求其可靠度。

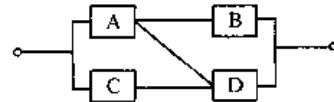


图 45.4-12 复杂系统例

**解** 现用两种方法来求:

1) 用全概率法,按式(45.4-19),取式中  $X$  为  $A$ ,则

$$R_s = P(A)P(S|A) + P(\bar{A})P(S|\bar{A})$$

当  $A$  未失效(犹如  $A$  处短路),则系统不失效的条件概率为  $B$  和  $D$  单元并联,故

$$\begin{aligned} P(A)P(S|A) &= R_A [1 - (1 - R_B)(1 - R_D)] \\ &= 0.9 [1 - (1 - 0.8)(1 - 0.6)] \end{aligned}$$

当  $A$  失效(犹如  $A$  处断路),则系统不失效的条件概率为  $C$  和  $D$  串联,故

$$P(\bar{A})P(S|\bar{A}) = (1 - R_A)R_C R_D$$

故该系统的可靠度

$$\begin{aligned} R_s &= R_A [1 - (1 - R_B)(1 - R_D)] + \\ &\quad R_C R_D (1 - R_A) \\ &= 0.9 [1 - (1 - 0.8)(1 - 0.6)] + \\ &\quad (1 - 0.9) \times 0.7 \times 0.6 = 0.8700 \end{aligned}$$

2) 用穷举列表法,因为单元数  $n=4$ ,故  $2^n=2^4=16$ ,将 16 种状态按全失效、3 个失效、2 个失效、1 个失效、全不失效列于表 45.4-1。按各行单元状态对图 45.4-12 判定系统状态。将系统各不失效状态的概率求出,即可求得  $R_s$ 。

$$R_s = \sum P_s = 0.8700$$

表 45.4-1 穷举表(布尔真值表)

系统状态序号	A $R_A=0.9$	B $R_B=0.8$	C $R_C=0.7$	D $R_D=0.6$	系统状态	$R_{ij}$
1	0	0	0	0	F	
2	0	0	0	1	F	
3	0	0	1	0	F	
4	0	0	1	1	S	$F_A F_B R_C R_D=0.0084$
5	0	1	0	0	F	
6	0	1	0	1	F	
7	0	1	1	0	F	
8	0	1	1	1	S	$F_A R_B R_C R_D=0.0336$
9	1	0	0	0	F	
10	1	0	0	1	S	$R_A F_B F_C R_D=0.0342$
11	1	0	1	0	F	
12	1	0	1	1	S	$R_A F_B R_C R_D=0.0756$
13	1	1	0	0	S	$R_A R_B F_C F_D=0.0864$
14	1	1	0	1	S	$R_A R_B F_C R_D=0.1296$
15	1	1	1	0	S	$R_A R_B R_C F_D=0.2016$
16	1	1	1	1	S	$R_A R_B R_C R_D=0.3024$

## 2 可靠性预计

可靠性预计也称可靠性预测，它是根据组成系统的各单元可靠性或以往经验来推测系统的可靠性。

### 2.1 可靠性预计的目的

- 1) 检验设计任务所提的可靠性指标是否能满足，预测产品的可靠度；
- 2) 进行方案比较，确定最优方案；
- 3) 查明可靠性薄弱环节；
- 4) 确认和验证可靠性增长；
- 5) 作为可靠性分配的基础；
- 6) 评价系统的固有可靠性。

### 2.2 可靠性预计的方法

可靠性预计有许多方法，随预计的目的、设计的时期、系统的规模、失效的类型及数据情况等的不同而用不同的方法。

#### 2.2.1 设计初期的概略预计法

新设计初期的预计，虽然没有足够的数，但对可行性研究、方案的比较等均起着重要的作用。这个方法的特点是：在统计了大量相似系统的性能参数与可靠性关系的基础上，进行分析，得出一些经验公式及系数，以便在方案论证及初步设计阶段，能根据初步确定系统性能及结构参数预计系统可靠性。

#### 2.2.2 数学模型法

数学模型法是可靠性预计所用的最主要方法。本方法根据各单元可靠性与系统可靠性关系所建立的精确或半精确的数学模型，通过计算预计系统的可靠性。

一般可仅考虑对系统可靠性有影响的主要组成，按可靠性的逻辑关系绘制可靠性框图，通常非串联部分均可单独计算，简化为一个等效单元，最终则成为一个简单的串联模型。故典型模型为

$$R_s(t) = \prod_{i=1}^n R_i(t) \quad (45.4-20)$$

式中  $R_i(t)$  —— 第  $i$  单元的可靠度， $i=1, 2, \dots, n$ 。

单元如是设备或装置等某一分系统，最好能有分系统的可靠性数据，否则需要将其分解成更小的单元，直到最基本的零件、元件。关于单元的可靠性数据可以运用以往积累的资料进行预计。资料来源于国家或企业的数据库、标准规范、参考资料及文献、外购件厂商数据、用户的调查、专门试验等。在设计中期和后期，则可按设计的详细资料对主要零部件或性能参数进行预计计算。

#### 2.2.3 上下限法

上下限法用于系统很复杂的情况，甚至由于考虑单元并不独立等原因不易建立可靠性预计的数学模型，就可用本方法预计得相当准确的预计值。对不太复杂的系统使用上下限法能比精确的数学模型法较快地求得预计值。本方法在绘得可靠性逻辑框图后，先考虑最简化的情况，再逐步复杂化，逐次算得系统可靠度的上限和下限，并在这上下限间取系统可靠度的预计值。现以图 45.4-13 为例介绍预计的方法。

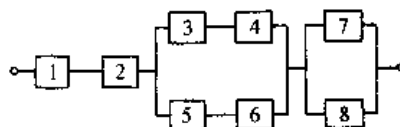


图 45.4-13 某系统可靠性逻辑图

(1) 上限计算



上限计算分两种情况来考虑:

1) 当系统中并联部分的可靠性很高时, 把它们  
的可靠度都近似取为 1。因此, 系统可靠度的上限只考虑  
串联单元, 系统可靠度上限一般形式为

$$R_{U0} = \prod_{i=1}^m R_i \quad (45.4-21)$$

式中  $R_i$  —— 第  $i$  个串联单元的可靠度;

$m$  —— 系统中的串联单元数。

2) 当系统中并联部分的可靠性不很高, 因而需  
计入并联单元对系统可靠度的影响。若仅考虑并联部分  
中仅有 1 个单元失效而引起系统失效的影响, 显然, 并  
联部分中单元 3、4、5、6、7、8 任一单元失效都不会  
引起系统失效, 即对系统无影响, 可靠度上限仍用式  
(45.4-21) 来求, 即  $R_{U1} = R_{U0}$ 。若仅考虑并联部分中  
有 2 个单元同时失效而引起系统失效的影响, 这时单  
元 3 和 5, 3 和 6, 4 和 5, 4 和 6, 7 和 8 中任一单元  
失效都会导致系统失效。当并联单元可靠度较高, 各对  
能引起系统失效的概率分别近似为:  $R_1 R_2 F_3 F_5$ ,  
 $R_1 R_2 F_3 F_6$ ,  $R_1 R_2 F_4 F_5$ ,  $R_1 R_2 F_4 F_6$ 。因此再考虑并联部  
分 2 个单元同时失效对系统失效的影响, 则系统可靠  
度的上限为类似, 系统可靠度上限一般形式为

$$R_{U2} = R_1 R_2 - R_1 R_2 (F_3 F_5 + F_3 F_6 + F_4 F_5 + F_4 F_6)$$

类似, 系统可靠度上限一般形式为

$$R_{U2} = \left( \prod_{i=1}^m R_i \right) \left( 1 - \sum_{(k,k')} F_k F_{k'} \right) \quad (45.4-22)$$

式中  $F_k, F_{k'}$  —— 分别为引起系统失效的一对单元各  
自的失效概率;

$p$  —— 2 个单元同时失效能引起系统失效  
的对数;

$R_i, m$  —— 同式 (45.5-19)。

(2) 下限计算

计算系统可靠度下限时, 首先是把系统中所有单  
元都算是串联, 即计算没有单元失效时系统不失效的  
概率, 故系统可靠度下限为

$$R_{L0} = \prod_{i=1}^n R_i \quad (45.4-23)$$

式中  $n$  —— 系统中所有单元的个数。

其实如果系统中任一单元失效, 例如图  
45.4-13 中单元 3、4、5、6、7、8 中任一失效, 并不  
会引起系统失效。这种状态的概率为

$$\begin{aligned} & R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 + R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 + \dots + R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 \\ & = R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 \left( \frac{F_3}{R_3} + \frac{F_4}{R_4} + \dots + \frac{F_8}{R_8} \right) \end{aligned}$$

若系统中并联单元数为 1, 则一般形式为

$$\left( \prod_{i=1}^n R_i \right) \left( \sum_{k=1}^l \frac{F_k}{R_k} \right)$$

这里,  $k$  表示失效并不引起系统失效的并联单元。

类似, 若并联部分有 2 个单元失效并不引起系统  
失效的对数为  $q$ , 这种状态的概率为

$$\left( \prod_{i=1}^n R_i \right) \left( \sum_{j,j'} \frac{F_j F_{j'}}{R_j R_{j'}} \right)$$

这里,  $j$  和  $j'$  表示同时失效但不引起系统失效的  
一对并联单元。

若取系统中没有单元失效和有一个并联单元失效  
而不引起系统失效的概率为系统可靠度下限, 则

$$R_{L1} = \left( \prod_{i=1}^n R_i \right) \left( 1 + \sum_{k=1}^l \frac{F_k}{R_k} \right) \quad (45.4-24)$$

式中  $R_k$  —— 第  $k$  个并联单元可靠度;

$F_k$  ——  $F_k \cdot 1 - R_k$ ;

$l$  —— 系统中并联单元数目。

其他符号同式 (45.4-23)。

若取系统中没有单元失效、有一个并联单元失效  
和二个并联单元失效而不引起系统失效的概率为系统  
可靠度下限, 则

$$\begin{aligned} R_{L2} = & \left( \prod_{i=1}^n R_i \right) \\ & \times \left( 1 + \sum_{k=1}^l \frac{F_k}{R_k} + \sum_{j,j'=1}^q \frac{F_j F_{j'}}{R_j R_{j'}} \right) \quad (45.4-25) \end{aligned}$$

式中  $R_j, R_{j'}$  —— 同时失效而不引起系统失效的一对  
并联单元的可靠度;

$F_j, F_{j'}$  ——  $F_j = 1 - R_j, F_{j'} = 1 - R_{j'}$ ;

$q$  —— 同时失效而不引起系统失效的并  
联单元对数;

其他符号同式 (45.4-24)。

(3) 组合预计

前面计算的上下限法随着计入并联或贮备部分同  
时失效数的增多而更趋逼近, 其计算也更趋复杂, 当计  
入同时失效数为  $s$ , 求得  $R_{U0} - R_{Ls} \approx 1 - R_{U0}$  时, 即可按  
下面组合公式预计系统的可靠度  $R_s$ 。

$$R_s = 1 - [(1 - R_{U0})(1 - R_{Ls})]^{1/2} \quad (45.4-26)$$

2.2.4 蒙特卡罗模拟法

当系统中各个单元的可靠性特征量已知, 但系统  
的可靠性模型过于复杂, 难以建立可靠性预计的精确  
数学模型或模型太复杂而不便应用则可用随机模拟法  
近似计算出系统可靠性的预计值。随着模拟次数的增  
多, 其预计精度也逐渐增高。由于需大量反复的计算,  
一般均用计算机来完成。

例 已知系统的可靠性逻辑框图如图 45.4-14 所示,  $R_1=0.92$ ,  $R_2=0.95$ ,  $R_3=0.94$ ,  $R_4=0.85$ ,  $R_5=0.9$ ,  $R_6=0.9$ ,  $R_7=0.88$ ,  $R_8=0.87$ ,  $R_9=0.86$ 。试用数学模型法和上下限法预计该系统的可靠度。

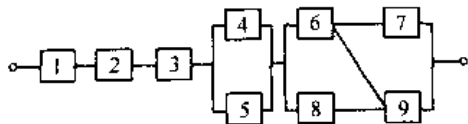


图 45.4-14 某系统可靠性逻辑框图

解 1) 用数学模型法 单元 4、5 为一并联系统, 用式 (45.4-5) 简化为一等效单元

$$R_{4,5} = 1 - (1 - R_4)(1 - R_5) \\ = 1 - (1 - 0.85)(1 - 0.9) = 0.985$$

单元 6、7、8、9 为一复杂系统

$$R_{6,7,8,9} = [1 - (1 - R_7)(1 - R_9)] \times \\ R_6 + R_8 R_9 (1 - R_6) \\ = [1 - (1 - 0.88)(1 - 0.86)] \times \\ 0.9 + 0.87 \times 0.86(1 - 0.9) = 0.9597$$

式 (45.4-20)

$$R_s = R_1 R_2 R_3 R_{4,5} R_{6,7,8,9} \\ = 0.92 \times 0.95 \times 0.94 \times 0.985 \times 0.9597 \\ = 0.7766$$

2) 用上下限法 按式 (45.4-21) 求上限

$$R_{U0} = R_{L1} = R_1 R_2 R_3 \\ = 0.92 \times 0.95 \times 0.94 = 0.8216$$

按式 (45.4-23) 求下限

$$R_{L10} = R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 R_9 \\ = 0.92 \times 0.95 \times 0.94 \times 0.85 \times 0.9 \times \\ 0.9 \times 0.88 \times 0.87 \times 0.86 = 0.3725$$

判断:  $1 - R_{L1} = 1 - 0.8216 = 0.1784$

$$R_{U0} - R_{L10} = 0.8216 - 0.3725 = 0.4491$$

相差颇大, 故再按式 (45.4-24) 求下限

$$R_{L1} = R_1 R_2 \cdots R_9 \left[ 1 + \frac{F_4}{R_4} + \frac{F_5}{R_5} + \cdots + \frac{F_9}{R_9} \right] \\ = 0.3725 \left[ 1 + \frac{0.15}{0.85} + \frac{0.1}{0.9} + \cdots + \frac{0.14}{0.86} \right] \\ = 0.6880$$

判断:  $R_{U0} - R_{L1} = 0.8216 - 0.6880 = 0.1336$  与  $1 - R_{L10} = 1 - R_{U0} = 0.1784$  相近, 故可按式 (45.4-26) 取

$$R_s = 1 - [(1 - R_{U0})(1 - R_{L1})]^{\frac{1}{2}} \\ = 1 - [(1 - 0.8216)(1 - 0.6880)]^{\frac{1}{2}} \\ = 0.7641$$

结果比数学模型法的精确预计值仅差约 1.6%。由于本例非串联单元可靠度不高, 故不宜再用式 (45.4-22) 求  $R_{U2}$ 。

### 3 可靠性分配

可靠性分配是将工程设计规定的系统可靠度指标合理地分配给组成该系统的各个单元, 并保证系统的可靠性指标得到满足。

可靠性分配本质上是一个工程决策问题。在进行可靠性分配时, 必须明确目标函数和约束条件。随目标函数和约束条件不同, 可靠性的分配方法也会有所不同。有的是以系统可靠度指标为约束条件, 把体积、重量、成本等系统参数尽可能小作为目标函数; 有的则给出体积、重量、成本等为约束条件, 要求将系统可靠度尽可能高地分配到各单元。此外, 还应根据系统的用途分析哪些参数应予以优先考虑, 哪些单元在系统中占有重要位置, 其可靠度应予以优先保证等来选择设计方案。

#### 3.1 可靠性分配的原则

通常分配应考虑下列原则:

1) 技术水平。对技术成熟的单元, 能够保证实现较高的可靠性, 或预期投入使用时可性可有把握地增长到较高水平, 则可分配给较高的可靠度。

2) 复杂程度。对较简单的单元, 组成该单元零部件数量少, 组装容易保证质量或故障后易于修复, 则可分配给较高的可靠度。

3) 重要程度。对重要的单元, 该单元失效将产生严重的后果, 或该单元失效常会导致全系统失效, 则应分配给较高的可靠度。

4) 任务情况。对整个任务时间内均需连续工作以及工作条件严酷, 难以保证很高可靠性的单元, 则应分配给较低的可靠度。

此外, 一般还要受费用、重量、尺寸等条件的约束。总之, 最终都是力求以最小的代价来达到系统可靠性的要求。

为了问题的简化, 一般均假定各单元的故障均互相独立。由于  $R=1-F$ , 对指数分布, 当  $F$  不大时,  $F \approx \lambda t$ , 因此可靠性分配可按情况将系统可靠度  $R$ , 分配给各  $i$  单元  $R_i$ , 当  $F_i$  很小时可将不可靠度  $F_i$  分配给各  $i$  单元  $F_i$ , 或者将系统的失效率  $\lambda$ , 分配给各  $i$  单元  $\lambda_i$ 。

#### 3.2 可靠性分配的方法

可靠性分配有许多方法, 随掌握可靠性资料的多少、设计的时期以及目标和限制条件等的不同而不同。下面仅介绍几种。

##### 3.2.1 等分配法

这是在设计初期, 当产品定义并不十分清晰时所

采用的最简单的分配方法，它是对系统中的全部单元分配以相等的可靠度的方法。

1) 串联系统 设系统由  $n$  个分系统串联组成，图 45.4-3 所示，若给定系统可靠度指标为  $R_s$ ，则按等分法，分配给各单元的可靠度指标  $R_i$  为

$$R_i = R_s^{\frac{1}{n}} \quad (45.4-27)$$

式中  $R_s$ ——系统要求的可靠度；  
 $R_i$ ——第  $i$  单元分配得的可靠度；  
 $n$ ——串联单元数。

2) 并联系统 如图 45.4-4 所示

$$F_i = F_s^{\frac{1}{n}} = (1 - R_i)^{\frac{1}{n}} \quad i = 1, 2, \dots, n \quad (45.4-28)$$

式中  $F_s$ ——系统要求的不可靠度；  
 $F_i$ ——第  $i$  单元分配得的不可靠度；  
 $R_s$ ——系统要求的可靠度；  
 $n$ ——并联单元数。

3) 混联系统 如图 45.4-15，一般可化为等效的单元，同级等效单元分配给相同的可靠度。例如图 45.4-15 中的单元可先按图 c，分配得

$$R_1 = R_{2,3,4} = R_s^{\frac{1}{2}}$$

再如图 b 分配得

$$R_2 = R_{3,4} = 1 - (1 - R_{2,3,4})^{\frac{1}{2}}$$

再求图 a 中的

$$R_3 = R_4 = R_{3,4}^{\frac{1}{2}}$$

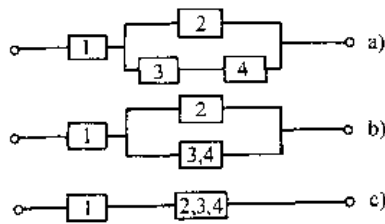


图 45.4-15 混联系统

### 3.2.2 再分配法

若通过预计知串联系统（可包括混联系统的等效单元）各单元的可靠度为  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_n$ 。则系统可靠度的预计值为

$$\hat{R}_s = \prod_{i=1}^n \hat{R}_i \quad i = 1, 2, \dots, n$$

若规定的系统可靠度  $R_s \leq \hat{R}_s$ ，表示预计值已满足规定的要求，各单元即可分配给  $\hat{R}_i$ 。反之，若  $R_s > \hat{R}_s$ ，表示预计值未满足规定的要求。需改进单元可靠度指标，即按规定的  $R_s$  指标进行再分配。由于提高低可靠

度单元的效果显著而且常较容易，因此只将低可靠度的单元按等分法进行再分配。为此先将各预计值按由小到大次序编号，则有

$$\hat{R}_1 \leq \hat{R}_2 \leq \dots \leq \hat{R}_m \leq \dots \leq \hat{R}_n$$

$$\text{令} \quad R_1 = R_2 = \dots = R_{m-1} = R_0$$

$$\text{当} \quad \hat{R}_m \leq R_0 = \left[ \frac{R_s}{\prod_{i=1}^{m-1} \hat{R}_i} \right]^{\frac{1}{m}} > \hat{R}_{m+1}$$

$$\text{则} \quad \left. \begin{aligned} R_1 = \dots = R_m = \left[ \frac{R_s}{\prod_{i=1}^m \hat{R}_i} \right]^{\frac{1}{m}} \\ R_{m+1} = \hat{R}_{m+1}, \dots, R_n = \hat{R}_n \end{aligned} \right\} (45.4-29)$$

用式 (45.4-29) 时由于  $m$  尚不知道，一般可暂设  $m$  进行试算。

例 已知四个单元串联系统各单元预计值由小到大分别为  $\hat{R}_1 = 0.9513, \hat{R}_2 = 0.9575, \hat{R}_3 = 0.9851, \hat{R}_4 = 0.9996$ 。若规定系统  $R_s = 0.95$ ，试进行可靠度再分配。

解 设  $m=1$ ，则

$$\begin{aligned} R_0 &= \left[ \frac{R_s}{\hat{R}_2 \hat{R}_3 \hat{R}_4} \right]^{\frac{1}{2}} \\ &= \left( \frac{0.95}{0.9757 \times 0.9851 \times 0.9996} \right)^{\frac{1}{2}} \\ &= 1.0076 > \hat{R}_2 \end{aligned}$$

不行，重设  $m=2$ ，则

$$\begin{aligned} R_0 &= \left[ \frac{R_s}{\hat{R}_3 \hat{R}_4} \right]^{\frac{1}{2}} = \left( \frac{0.95}{0.9851 \times 0.9996} \right)^{\frac{1}{2}} \\ &= 0.9822 > \hat{R}_1 \text{ 行} \\ &> \hat{R}_2 \end{aligned}$$

故取  $R_1 = R_2 = 0.9822$

$$R_3 = \hat{R}_3 = 0.9851, R_4 = \hat{R}_4 = 0.9996$$

### 3.2.3 比例分配法

本法用于新设计的系统与原有系统基本相同，已知原有系统各单元  $\hat{F}_i$  或失效率  $\hat{\lambda}_i$ ，但对新设计的系统规定了新的可靠性要求，或者根据已掌握的可靠性资料，已能预计新设计系统各单元的  $\hat{F}_i$  或  $\hat{\lambda}_i$ ，但尚未满足新设计系统可靠性的要求。这时可取新系统分配给各单元的失效概率  $\hat{F}_i$  与相应单元  $\hat{\lambda}_i$  成正比，若为指数分布，则各单元分配的失效率  $\lambda_i$  与相应单元  $\hat{\lambda}_i$  成正比。

1) 串联系统 若系统要求可靠度为  $R_s$ ，

$$F_i \approx \frac{F_s \hat{F}_i}{\sum_{i=1}^n \hat{F}_i} = \frac{(1 - R_s) \hat{F}_i}{\sum_{i=1}^n \hat{F}_i} \quad (45.4-30)$$

当各单元寿命服从指数分布

$$\lambda_i = \frac{\lambda_s \hat{\lambda}_i}{\sum_{i=1}^n \hat{F}_i} \quad (45.4-31)$$

例1 已知某系统为4个单元串联,原系统工作100h各单元失效概率分别为  $\hat{F}_1 = 0.0425$ ,  $\hat{F}_2 = 0.0149$ ,  $\hat{F}_3 = 0.0487$ ,  $\hat{F}_4 = 0.0004$ ,新设计要求工作100h的可靠度  $R_s = 0.95$ ,求分配给各单元的可靠度。

解 用式(45.4-30),式中

$$\begin{aligned} \sum_{i=1}^4 \hat{F}_i &= 0.0425 + 0.0149 + 0.0487 + 0.0004 \\ &= 0.1065 \end{aligned}$$

$$F_s = 1 - R_s = 1 - 0.95 = 0.05$$

$$F_1 = \frac{0.05}{0.1065} \times 0.0425 = 0.01995$$

$$R_1 = 1 - F_1 = 0.98005$$

$$F_2 = \frac{0.05}{0.1065} \times 0.0149 = 0.0070$$

$$R_2 = 1 - F_2 = 0.9930$$

$$F_3 = \frac{0.05}{0.1065} \times 0.0487 = 0.0229$$

$$R_3 = 1 - F_3 = 0.9771$$

$$F_4 = \frac{0.05}{0.1065} \times 0.0004 = 0.00019$$

$$R_4 = 1 - F_4 = 0.99981$$

验算

$$\begin{aligned} R_s &= R_1 R_2 R_3 R_4 \\ &= 0.98005 \times 0.9930 \times 0.9771 \times 0.99981 \\ &= 0.9507 \approx 0.95 \end{aligned}$$

例2 已知前例各单元寿命为指数分布,失效率分别为  $\hat{\lambda}_1 = 0.000425$  1/h,  $\hat{\lambda}_2 = 0.000149$  1/h,  $\hat{\lambda}_3 = 0.000487$  1/h,  $\hat{\lambda}_4 = 0.000004$  1/h,新设计要求  $R_s = 0.95$ ,求各单元应有的可靠度。

解 用式(45.4-31),式中

$$\begin{aligned} \lambda_s &= \frac{1}{t} \left( \ln \frac{1}{R_s} \right) = \frac{1}{100} \ln \frac{1}{0.95} \\ &= 0.0005129 \quad 1/h \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^4 \hat{\lambda}_i &= 0.000425 + 0.000149 + 0.000487 + 0.000004 \\ &= 0.001065 \quad 1/h \end{aligned}$$

$$\lambda_1 = \frac{0.0005129}{0.001065} \times 0.000425$$

$$= 0.0002047 \quad 1/h$$

$$\begin{aligned} \lambda_2 &= \frac{0.0005129}{0.001065} \times 0.000149 \\ &= 0.00007176 \quad 1/h \end{aligned}$$

$$\begin{aligned} \lambda_3 &= \frac{0.0005129}{0.001065} \times 0.000487 \\ &= 0.00002345 \quad 1/h \end{aligned}$$

$$\begin{aligned} \lambda_4 &= \frac{0.0005129}{0.001065} \times 0.000004 \\ &= 0.00000193 \quad 1/h \end{aligned}$$

各单元应分配的可靠度

$$R_1 = e^{-\lambda_1 t} = e^{-0.0002047 \times 100} = 0.97974$$

$$R_2 = e^{-\lambda_2 t} = e^{-0.00007176 \times 100} = 0.9928$$

$$R_3 = e^{-\lambda_3 t} = e^{-0.00002345 \times 100} = 0.9768$$

$$R_4 = e^{-\lambda_4 t} = e^{-0.00000193 \times 100} = 0.9998$$

验算

$$\begin{aligned} R_s &= R_1 R_2 R_3 R_4 = 0.97974 \times \\ &0.9928 \times 0.9768 \times 0.9998 \\ &= 0.9499 \approx 0.95 \end{aligned}$$

2) 并联系统 若系统要求不可靠度为  $F_s$ ,

$$F_i = \left[ \frac{F_s}{\prod_{i=1}^n F_i} \right]^{\frac{1}{n}} \hat{F}_i \quad (45.4-32)$$

当各单元寿命服从指数分布时

$$\lambda_i = \left[ \frac{F_s}{\prod_{i=1}^n \hat{F}_i} \right]^{\frac{1}{n}} \frac{\hat{\lambda}_i}{t} \quad (45.4-33)$$

例3 已知某系统为3个单元并联,预计得工作1000h各单元失效概率分别为  $\hat{F}_1 = 0.08$ ,  $\hat{F}_2 = 0.10$ ,  $\hat{F}_3 = 0.15$ ,新设计要求工作100h时  $R_s = 0.9995$ ,求各单元应分配的可靠度。

解 用式(45.4-32),式中

$$\begin{aligned} \prod_{i=1}^3 \hat{F}_i &= \hat{F}_1 \hat{F}_2 \hat{F}_3 = 0.08 \times 0.10 \times 0.15 \\ &= 0.0012 \end{aligned}$$

$$F_s = 1 - R_s = 1 - 0.9995 = 0.0005$$

$$F_1 = \left( \frac{0.0005}{0.0012} \right)^{\frac{1}{3}} \times 0.08 = 0.0598$$

$$R_1 = 1 - F_1 = 1 - 0.0598 = 0.9402$$

$$F_2 = \left( \frac{0.0005}{0.0012} \right)^{\frac{1}{3}} \times 0.10 = 0.0747$$

$$R_2 = 1 - F_2 = 1 - 0.0747 = 0.9253$$

$$F_3 = \left( \frac{0.0005}{0.0012} \right)^{\frac{1}{3}} \times 0.15 = 0.1120$$

$$R_3 = 1 - F_3 = 1 - 0.112 = 0.8880$$

验算:

$$R_s = 1 - F_1 F_2 F_3 = 1 - 0.0598 \times 0.0747 \times 0.112 = 0.9995$$

例 4 若前题各单元寿命为指数分布, 预计得各单元失效率  $\lambda_1 = 0.00008$  1/h,  $\lambda_2 = 0.00010$  1/h,  $\lambda_3 = 0.00015$  1/h, 新设计要求  $R_s = 0.9995$ , 求各单元应分配的可靠度。

解 用式 (45.4-33), 式中

$$\prod_{j=1}^3 \lambda_j = \lambda_1 \lambda_2 \lambda_3 = 0.00008 \times 0.0001 \times 0.00015$$

$$= 1.2 \times 10^{-12} (1/h)^3$$

$$F_s = 1 - R_s = 1 - 0.9995 = 0.0005$$

$$\lambda_1 = \left( \frac{0.0005}{1.2 \times 10^{-12}} \right)^{\frac{1}{3}} \frac{0.00008}{1000}$$

$$= 0.00005975 \quad 1/h$$

$$\lambda_2 = \left( \frac{0.0005}{1.2 \times 10^{-12}} \right)^{\frac{1}{3}} \frac{0.0001}{1000}$$

$$= 0.00007469 \quad 1/h$$

$$\lambda_3 = \left( \frac{0.0005}{1.2 \times 10^{-12}} \right)^{\frac{1}{3}} \frac{0.00015}{1000}$$

$$= 0.0001120 \quad 1/h$$

$$R_1 = e^{-\lambda_1 t} = e^{-0.00005975 \times 10^6} = 0.942$$

$$R_2 = e^{-\lambda_2 t} = e^{-0.00007469 \times 10^6} = 0.928$$

$$R_3 = e^{-\lambda_3 t} = e^{-0.000112 \times 10^6} = 0.894$$

验算

$$R_s = 1 - (1 - R_1)(1 - R_2)(1 - R_3)$$

$$= 1 - (1 - 0.942)(1 - 0.928)(1 - 0.894)$$

$$= 0.9996 \approx 0.9995$$

3) 混联系统 化为等效单元后, 分别运用比例分配法。

### 3.2.4 综合评分分配法 (AGREE)

本分配法是按经验对各单元考虑主要因素综合评

分, 根据各单元得分多少分配给相应的可靠性指标。关于考虑的因素, 评分办法也可视具体情况而定, 高分则分给较高的失效概率或失效率。一般, 考虑的因素包括: ①技术水平: 对技术成熟, 有把握保证高可靠性评 1 分, 反之评 10 分; ②复杂程度: 单元组成元件少, 结构简单评 1 分, 反之评 10 分; ③重要程度: 极其重要评 1 分, 反之评 10 分; ④任务情况: 整个任务期中工作时间很短, 工作条件好评 1 分, 后之评 10 分。第  $i$  单元综合得分可取各因数得分之积, 即

$$\omega_i = \prod_{j=1}^4 \omega_{ij} \quad (45.4-34)$$

式中  $j=1, 2, 3, 4$ , 分别代表前述四项因素系统总分数。

$$\omega = \sum_{i=1}^n \omega_i \quad (45.4-35)$$

式中  $i=1, 2, \dots, n$  为单元编号。

第  $i$  单元得分比数

$$\epsilon_i = \frac{\omega_i}{\omega} \quad (45.4-36)$$

一般串联系统分配得

$$R_i = R_s^{\epsilon_i} \quad (45.4-37)$$

当各单元寿命为指数分布, 则

$$\lambda_i = \epsilon_i \lambda_s = \frac{\epsilon_i}{t} \ln \frac{1}{R_s} \quad (45.4-38)$$

$$R_i = e^{-\lambda_i t} = R_s^{\epsilon_i} \quad (45.4-39)$$

例 某系统由 4 个单元串联, 要求任务时间为 100h 的可靠度  $R_s = 0.90$ , 按综合评分法求各单元的可靠度。若各单元为指数分布, 求各单元应分配的失效率。

解 用式 (45.4-37、38) 列表计算 (见表 45.4-2)。

表 45.4-2 计算结果

单元号 $i$	评 分				单元积分 $\omega_i = \prod_{j=1}^4 \omega_{ij}$	单元得分比 $\epsilon_i = \frac{\omega_i}{\omega}$	分得可靠度 $R_i = R_s^{\epsilon_i}$	分得失效率 $\lambda_i = \frac{\epsilon_i}{100} \ln \frac{1}{0.9} (1/h)$
	技术水平 $\omega_{i1}$	复杂程度 $\omega_{i2}$	重要程度 $\omega_{i3}$	任务情况 $\omega_{i4}$				
1	2	2	5	5	100	0.00785	0.9992	$8.27 \times 10^{-6}$
2	6	4	7	10	1680	0.13187	0.9862	$1.39 \times 10^{-4}$
3	10	10	10	10	10000	0.78493	0.9206	$8.27 \times 10^{-4}$
4	8	3	8	5	960	0.07535	0.9921	$7.94 \times 10^{-5}$
$\Sigma \omega_i = 12740$						$R_s = 0.90$		$\lambda_s = 0.001054$

### 3.2.5 动态规划分配法

本方法是解决在满足规定的系统可靠性指标的条件下, 使费用或重量, 或者尺寸等最小的优化问题。最常用是使费用最小, 下面即以最小费用为例。

(1) 串联系统

$$\text{目标函数 } \min \sum_{i=1}^n G_i(\hat{R}_i, R_i)$$

$$\text{约束条件 } \prod_{i=1}^n R_i \geq R_s$$

$$\text{和 } 0 < \hat{R}_i \leq R_i \leq 1 \quad i = 1, 2, \dots, n$$

(45.4-40)

式中  $R_s$ ——规定的系统可靠度；  
 $R_i$ ——第  $i$  单元分配得的可靠度；  
 $\hat{R}_i$ ——第  $i$  单元现有可靠度；  
 $G_i(\hat{R}_i, R_i)$ ——第  $i$  单元可靠度由  $\hat{R}_i$  提高到  $R_i$  的所需费用函数。

下面通过例题来说明利用动态规划作可靠度分配的方法。

例1 某系统为3个单元串联,要求系统可靠度  $R_s \geq 0.96$ 。按现在已掌握的资料预计得各单元可靠度为  $\hat{R}_1 = 0.95, \hat{R}_2 = 0.96, \hat{R}_3 = 0.98$ , 费用函数如表

表 45.4.3 费用函数

$R_1$	$G_1(0.95, R_1)$	$R_2$	$G_2(0.96, R_2)$	$R_3$	$G_3(0.98, R_3)$
0.96	1.0	0.96	0		
0.97	2.0	0.97	2.0		
0.98	4.0	0.98	5.0	0.98	0
0.99	12.0	0.99	15.0	0.99	8
0.995	50.0	0.995	35.0	0.995	20

表 45.4.4  $R_1$  和  $R_2$  组合方案列表计算

$R_1(G_1) \backslash R_2(G_2)$	0.97(2)	0.98(5)	0.99(15)	0.995(15)
0.97(2)	0.9409	0.9506	0.9603(17)	0.96515(37)
0.98(4)	0.9506	0.9604(9)	0.9702(19)	0.9751(39)
0.99(12)	0.9603(14)	0.9702(17)	0.9801(27)	0.98505(47)
0.995(50)	0.96515(52)	0.9751(55)	0.98505(65)	0.99002(85)

表 45.4.5  $R_1, R_2$  和  $R_3$  组合方案列表计算

$R_3(G_3) \backslash R_1 R_2(G_1 \cdot G_2)$	0.9604(9)	0.9702(17)	0.9801(27)	0.98505(47)	0.99002(85)
0.98(0)	0.9415(9)	0.9508(17)	0.9605(27)	0.96535(47)	0.9702(85)
0.99(8)	0.9511(17)	* 0.9605(25)	0.9703(35)	0.9752(55)	0.9801(93)
0.995(20)	0.9559(29)	0.9653(37)	0.9752(47)	0.9801(67)	0.9851(105)

(2) 并联系统

$$\left. \begin{aligned} \text{目标函数 } & \min \sum_{i=1}^n G_i(\hat{F}_i, F_i) \\ \text{约束条件 } & \prod_{i=1}^n F_i \leq F_s \\ & 0 < F_i \leq \hat{F}_i < 1 \\ & i = 1, 2, \dots, n \end{aligned} \right\} (45.4-41)$$

式中  $F_s$ ——规定的系统不可靠度；  
 $F_i$ ——第  $i$  单元分配得的不可靠度；  
 $\hat{F}_i$ ——第  $i$  单元现有不可靠度；  
 $G_i(\hat{F}_i, F_i)$ ——第  $i$  单元不可靠度由  $\hat{F}_i$  降到  $F_i$  的所需费用函数。

下面举例说明分配法。

例2 某系统为3个单元并联,要求系统可靠度  $R_s \geq 0.9995$ 。按现在掌握的资料预计得各单元不可靠度  $\hat{F}_1 = 0.10, \hat{F}_2 = 0.10, \hat{F}_3 = 0.12$ , 费用函数如表

45.4-3 所示。为使总费用最小,求各单元应分配的可靠度。

解

先将  $R_1$  和  $R_2$  的组合方案列表计算于表 45.4-4,  $R_i \leq R_s$  的项不必列入。

将表 45.4-4 中  $R_1 R_2 > R_s$ , 且费用相对最小的组合值再与  $R_3$  组合列表计算于表 45.4-5。

表 45.4-5 中“\*”的方案  $\prod_{i=1}^n R_i > R_s$ , 且费用最小。由此可知,应取  $R_1 = 0.99, R_2 = 0.98, R_3 = 0.99$ 。

45.4-6 所示。为使费用最小,求各单元应分配的不可靠度。

表 45.4.6 费用函数

$F_1$	$G_1(0.1, F_1)$	$F_2$	$G_2(0.1, F_2)$	$F_3$	$G_3(0.12, F_3)$
0.10	0	0.10	0	0.10	1
0.08	2	0.08	3	0.08	4
0.06	6	0.06	5	0.06	12
0.04	15	0.04	13	0.04	20

解 1) 算出规定的总失效概率  $F_s \leq 1 - 0.9995 = 0.0005$ 。

2) 先将  $F_1$  和  $F_2$  组合方案列表计算于表 45.4-7。

3) 将表 45.4-8 中  $F_1 F_2 (G_1 + G_2)$  中  $F_1 F_2$  和  $G_1 + G_2$  中小的组合再与  $F_3 (G_3)$  组合列表计算于表 45.4-8, 显然,  $F_1 F_2 \leq 0.005$  中  $G_1 + G_2$  显著大的也不必列入。

4) 表 45.4-8 中注“\*”者为满足  $F_s \leq 0.0005$  费用最低的组合,其他未计算者费用都高于 8。由此可知,应取  $F_1 = 0.08, F_2 = 0.06, F_3 = 0.10$ 。

表 45.4-7  $F_1$  和  $F_2$  组合方案列表计算

$F_1(G_1) \backslash F_2(G_2)$	0.10(0)	0.08(3)	0.06(5)	0.04(13)
0.10(0)	0.01(0)*	0.008(3)	0.006(5)*	0.004(13)
0.08(2)	0.008(2)*	0.0064(5)	0.0048(7)*	0.0032(15)
0.06(6)	0.006(6)	0.0048(9)	0.0036(11)*	
0.04(15)	0.004(15)	0.0032(18)		

表 45.4-8  $F_1, F_2$  和  $F_3$  组合方案列表计算

$F_1 F_2(G_1 \cdot G_2) \backslash F_3(G_3)$	0.12(0)	0.10(1)	0.08(4)	0.06(12)	0.04(20)
0.01(0)	0.0012	0.001	0.0008	0.0006	0.0004(20)
0.008(2)	0.00096	0.0008	0.00064	0.00048(14)	
0.006(5)	0.00072	0.0006	0.00048(9)	0.00036(17)	
0.0048(7)	0.00058	*0.00048(8)	0.00038(11)		
0.0036(11)	0.000432(11)				

## 4 失效模式、效应及危害度分析 (FMECA)

### 4.1 基本概念

失效模式影响及危害性分析 (Failure Mode Effect and Criticality Analysis, 简称 FMECA) 定义为“在系统设计过程中, 通过对系统各组成单元潜在的各种失效模式及其对系统功能的影响与产生后果的严重程度进行分析, 提出可能采取的预防改进措施, 以提高产品可靠性的一种设计分析方法”(GB/T3187—1994)。失效模式是失效表现的状态。失效效应是指某种失效模式对本单元和整个系统的影响, 失效危害度是指失效后果的危害程度。当只进行故障模式和影响分析时, 简称为 FMEA。

FMECA 是一种自下而上的故障分析方法, 即以分析零件故障产生的后果及危害程度为基础来分析部件的故障, 由部件故障进一步分析整个系统或整机的故障。它可以应用于产品开发、生产和使用的不同阶段, 应用于设计评审、可靠性试验分析、工艺分析、使用故障分析等。

本方法既是协助设计评审的一种技术, 也是一种保证及评价方法, 在系统和子系统设计的最初阶段就可加以应用, 并在设计过程中根据获得的新资料不断改进。它可用于各种级别的系统设计, 是提高产品可靠性的一种设计分析方法。

### 4.2 分析的过程与方法

由于产品的多样性, 分析过程和方法也不尽相同, 大致可取下列内容适当增减。

#### (1) 划分功能块

系统可逐级分解直到最基本的零件、构件。一般根

据分析的目的, 可仅将系统分解到某一水平。将系统按功能分解为功能块并绘出系统功能逻辑框图。

#### (2) 列举各功能块所有失效模式、起因和效应

失效模式应与该功能块所在级相适应。在最低的分析级上, 列出该级各单元所有可能出现的各种失效模式, 以及每种失效模式发生的起因、对应的失效效应。对于一个更高功能级上考虑失效效应时, 前述失效效应又被解释为一个失效模式。连续迭代直至系统最高功能级上的失效效应。如表 45.4-9 所示阀门失效模式与原因表。为避免重大遗漏, 宜由熟悉该系统结构、工作原理、使用情况的设计、使用等技术人员共同分析。

表 45.4-9 阀门故障模式与原因表

故障模式	故障原因
内部泄漏	阀体、阀座变形, 损伤 阀体、阀座接触面有异物
外部泄漏	密封部损伤
破损	长期使用后疲劳破视 腐蚀 外力
堵塞	进入异物 阀杆断, 阀体落下
误关/误开	误操作 误信号
不关/不开	异物阻碍 驱动装置(电动机, 传动机构等)故障 丧失动力(电、压缩空气等)
不能控制	控制零件(如弹簧)失效

#### (3) 失效等级与危害度分析

FMEA 中的失效等级也称为重要度, 是反映失效模式重要程度的综合指标。通常都是采用相对评分法决定其等级, 如以完成任务为重点的评分法, 以故障发生频度为重点的评分法和综合考虑多种因素的综合评分法。下面介绍综合评分法。表 45.4-10 列出常用故障

模式的评定因素及其评分范围,按此表逐项评分,然后按式(45.5-42)计算危害度系数(致命度系数) $C_i$ 。 $C_i$ 值越高,失效模式的危害度越高。

$$C_i = \prod_{j=1}^n F_j \quad (45.4-42)$$

式中  $F_i$ ——第  $i$  项评定因数的评分值;

$n$ ——考虑评定因素的项数。

(4) 提出改进措施

应尽量消除危害性高的失效模式,这时常需改变设计。无法消除时,应分配给高的可靠性指标,必要时,增设报警、监测、防护等设施。

(5) 填写失效模式、效应及危害度分析的表格

不同系统所用表格不尽相同,但对同类产品,企业内部宜取统一格式。图 45.4-16 为典型表格的例子。

表 45.4-10 故障等级与致命度系数评分

	综合评分法 评分 $C_i$	致命度系数法	
		程 度	系数 $F_i$
1. 故障对功能的影响及后果	1~10	致命的损失	5.0
		相当大的损失	3.0
		丧失功能	1.0
		不丧失功能	0.5
2. 故障对系统的影响范围	1~10	两个以上重大影响	2.0
		一个重大影响	1.0
		无太大影响	0.5
3. 故障发生频度	1~10	发生频度高	1.5
		有发生的可能性	1.0
		发生的可能性很小	0.7
4. 故障防止的可能性	1~10	不能防止	1.3
		可能防止	1.0
		可容易地防止	0.7
5. 更改设计的程度		须作重大改变	1.2
		须作类似设计	1.0
		同一设计	0.8

失效模式效应及危害度分析								第 页
系统 _____		填表人 _____		日期 _____				
方块图号	失效模式	判定原因	效 应		检测方法	危害度	改进措施	备注
			本单元	系统				
			(发生后果)			(评等级或评分)		

失效模式效应及危害度分析										第 页
系统 _____		填表人 _____		日期 _____						
零件名称 零件号	零件功用	故障模式	故障效应	故障原因	发生频率程度	影响严重程度	发现难易程度	危害顺序数 RPN	改进措施和 目前情况	
					$N_f$	$N_e$	$N_d$			
		(列举)			(1—10分)	(1—10分)	(1—10分)	前三项分乘积		

图 45.4-16 失效模式、效应及危害度分析表格例

4.3 FMECA 应用例子

对某拖拉机液压系统进行 FMECA 分析,液压系统如图 45.4-17。液压泵由 4 个对称卧式柱塞分泵组成,液压泵偏心轴上有两个互成 90° 偏心轮,偏心轮被套在外方内圆的衬套内,衬套嵌在柱塞架内。每个柱塞

架上有两个对称的柱塞,当偏心轴转动时,带动柱塞在柱塞缸内往复运动,完成吸油或压油功能。使阀体内腔分隔成进油室及回油室,由它来控制油液进出液压泵。控制阀有四种工作位置:平衡位、进油位、快速回油位及慢速回油位。在油路中装有球型安全阀,安全阀的开启压力由压力弹簧控制。当安全阀全开时,来自液压泵



的全部高压油都可以通过安全阀口排出，回到主液压缸内，使系统压力不再增加。防止系统压力超载，以免造成机件损坏。油液推动活塞，带动悬挂机构提升农具。当操纵液压系统的回油路打开时，在农具自重的作用下，油液被活塞压出液压缸使农具下降。来自液压泵的高压油，也可进行液压输出。

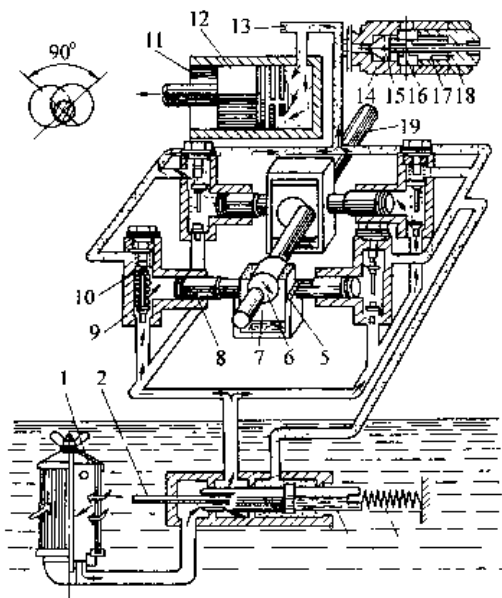


图 45.4-17 液压系统油路图

- 1—滤网总成 2—阀杆 3—控制阀 4—弹簧
- 5—柱塞架 6—偏心轮 7—宽偏心轴衬 8—柱塞
- 9—进油阀 10—出油阀 11—液压缸 12—活塞
- 13—液压输出螺塞 14—安全阀体 15—阀杆
- 16—弹簧 17—调节螺丝 18—螺母 19—偏心轮及轴

FMECA 分析的具体步骤如下：

1) 弄清所分析系统零件构成及功能，用框图绘出分析，如图 45.4-18 所示。在确定分析范围时，为简化起见，将故障发生频度低，对系统故障影响小的零件，未列入分析范围之内。

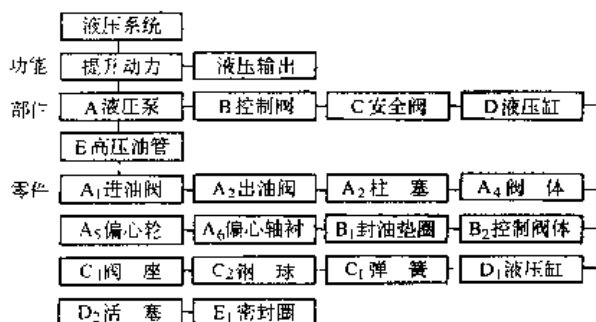


图 45.4-18 液压系统分析范围及等级框图

2) 列出所分析范围内的主要零部件可能出现故障模式，并分析其原因。如表 45.4-11。

在填写 FMECA 分析表时，应注意以下几点：

①应根据系统所有零部件可能出现的故障模式，确定分析范围，对重要的子系统和部件应进行重点分析。②根据零件出现故障严重程度、发生故障频度及发现和查明故障的难易程度，确定综合评定指标。根据该指标，确定改进的重点或先后顺序。③在表格中要写明改进措施、负责实施的部门及完成期限。④在编制此表时应尽可能汇总设计、工艺、制造、试验、使用维修等部门人员的经验和知识，对每种故障模式进行全面的分析。

表 45.4-11 某拖拉机液压系统 FMECA 分析表

零部件名称	故障模式	功能	故障后果	故障原因	影响度			综合评定指标	改进措施
					频度 $F_1$	严重程度 $F_2$	难易度 $F_3$		
液 压 泵	进油阀 渗油	进油或阻止高压油回流	一部分高压油倒流回后桥箱体内注；液压泵置于后桥箱内	加工误差或使用磨损，均会造成阀体与阀座之间间隙过大	2	2	3	12	1. 制造部门应保证加工精度 2. 使用中要保持油液清洁
	出油阀 渗油	柱塞往复运动造成柱塞缸内压力差，使出油阀完成出油或阻止高压油回流	使一部分高压油倒流入柱塞缸内	加工误差或使用磨损，均会造成阀体与阀座之间间隙过大	2	2	3	12	1. 制造部门应保证加工精度 2. 使用中要保持油液清洁
	柱塞与缸体 泄漏	柱塞在缸体内作往复运动，完成吸油与压油工作过程	压油量减少，输出压力降低	加工误差或使用磨损，造成柱塞与缸体之间间隙过大	1	3	3	9	1. 制造部门应保证柱塞与缸体加工精度 2. 使用中要保持油液清洁
偏心轮与柱塞架	卡死	偏心轮带动柱塞架作往复运动	柱塞架碎裂使柱塞不能完成吸油和压油	偏心轮与衬套加工误差，造成两者配合过紧，尤其是在重载荷时，油温过高，易出现偏心轮在衬套内卡死造成柱塞架碎裂	3	8	4	96*	1. 制造部门应保证偏心轮与衬套加工精度，使其间隙符合设计要求 2. 设计部门应改进柱塞架材料及加工工艺，提高其强度

(续)

零部件名称	故障模式	功能	故障后果	故障原因	影响度			综合评定指标	改进措施
					频度 $F_1$	严重度 $F_2$	难易度 $F_3$		
偏心轴衬	磨损快	偏心轮通过偏心轴衬, 带动柱塞架往复运动	由于偏心轴衬磨损, 使活塞行程缩短、压油量减少	偏心轴衬为铜衬套, 材质及加工达不到设计要求时, 磨损过快, 长期使用也会使轴衬磨损量增大	2	3	4	24	1. 制造部门应保证偏心轴衬材料及加工工艺 2. 设计部门改进偏心轴衬材料, 提高其耐磨性
控制阀	封油垫圈漏油	三片封油垫圈分隔阀体内腔为进油室及回油室, 控制油液出入液压泵	控制阀处于平衡位置时, 封油垫圈漏油, 使液压系统内泄量增大, 静沉降值增大	1. 由于封油垫圈与阀体加工误差造成配合间隙过大 2. 封油垫圈较薄, 易磨损泄漏	4	8	4	128*	设计部门应改进设计, 采用改进结构的控制阀
安全阀	阀座与钢球	钢球被弹簧压入阀座内, 控制系统压力, 防止压力过载	安全阀尚未开启时, 由于钢球与阀座间隙大, 造成系统内压力油泄漏, 提升系统提升能力降低	1. 由于阀座与钢球加工质量造成两者结合面间隙过大 2. 长期使用, 使钢球及阀座磨损	2	3	2	12	制造部门应保证阀座与钢球加工精度及配合间隙, 检验部门应严格检验安全阀在开启之前的渗油情况, 保证密封性能
	弹簧	水久变形量大	控制安全阀的开启压力及全开压力	导致安全阀的开启压力及全开压力降低, 使系统提升能力下降, 静沉降值增大	1. 出厂时安全阀开启压力调整偏低 2. 弹簧受力后水久变形量大	4	7	3	84*
液压缸	缸体与活塞	来自液压泵的高压油, 通过液压缸内推动活塞, 带动悬挂机构提升	缸体与活塞间渗漏, 造成系统内泄量增加, 静沉降值增大, 提升能力下降	柱塞与缸体的加工质量未达到要求, 造成两者配合间隙过大	1	2	2	4	制造部门应保证活塞与缸体加工质量及配合间隙
高压油管	密封圈	高压管是连接液压泵和提升机构的油道	密封圈损坏, 造成液压系统漏油, 提升能力下降或不能提升	高压油管上的密封圈由于装拆不当, 致使密封圈损坏	2	3	1	6	装配及修理部门在拆装高压油管时, 要防止密封圈损坏

3) 根据故障发生频度  $F_1$ , 故障危害程度  $F_2$  及故障发现和查明的难易程度  $F_3$  来确定综合评定指标  $F$

$$F = F_1 F_2 F_3$$

$F_1$ 、 $F_2$  和  $F_3$  的推荐值见表 45.4-12。

表 45.4-12  $F_1$ 、 $F_2$ 、 $F_3$  推荐值

$F_1$ (故障发生频度)			$F_3$ (故障发现和查明的难易程度)		
频度等级	判 据	系 数	难易度等级	判 据	系 数
I	>5%~20%	5	I	很难发现和查明的故障	5
II	>1%~5%	3~4	II	难以发现和查明的故障	3~4
III	>0.3%~1%	2	III	较难发现和查明的故障	2
IV	≤0.3%	1	IV	容易发现和查明的故障	1

$F_2$ (故障严重程度)			
严重度等级	名称及代号	判 据	系 数
I	致命故障 ZM	按各类故障定义判别	9~8
II	严重故障 YZ		6~8
III	一般故障 YB		3~5
IV	轻度故障 QD		1~2

4) 根据步骤 2) 提出改进措施; 根据步骤 3) 确定改进重点及先后顺序。

从表 45.4-11 中看出, 要提高液压系统的性能及可靠性, 应重点解决综合评定指标值高的三个问题:

a) 偏心轮与柱塞架卡死, 造成柱塞架断裂的故障; b) 控制阀封油垫圈渗漏, 造成液压系统内泄量增大, 静沉

降值增加；c) 安全阀弹簧在使用中永久变形量大，使安全阀开启压力及全开压力下降，导致液压系统提升能力下降。

该机液压系统经过 FMECA 分析，提出了解决的措施：柱塞架采用新材料，提高其韧性，控制阀进行结构改进，主要解决控制阀封油垫圈漏油问题。加大了封油垫圈的厚度，安全阀参考国外引进技术，在结构上进行改进，这些改进措施采用后，取得较好效果。

改进设计采用整体式结构控制阀，由阀套和阀杆两个零件构成，替代了原结构中的九个零件（压套、衬套、阀杆及挡圈各 1 件、垫圈 2 只、封油垫圈 3 只）。装有改进控制阀的液压泵（以下称改进泵）与原泵进行了性能对比试验、台架寿命试验及田间使用试验。试验结果表明：改进泵内部泄漏量显著减小。尤其是在长期使用中，阀套及阀杆的磨损极小，达到了液压系统静沉降值规定的要求，提高了液压泵的性能及可靠性。

## 5 故障树分析 (FTA)

### 5.1 基本概念

故障树分析也叫失效树分析，故障树分析 (Fault Tree Analysis 简称 FTA) 是“在系统设计过程中，通过对造成系统失效的各种因素(包括硬件、软件、环境、人为因素)进行分析，画出逻辑框图(即故障树)，从而确定系统失效原因的各种可能组合方式及其发生的逻辑关系，以计算系统失效概率，采取相应的纠正措施，以提高系统可靠性的一种设计分析方法”(GB/T3187—1994)。其特点为：

1) 故障树分析法是一种图形演绎方法，是故障事件在一定条件下的逻辑推理方法，它不局限于对系统作一般的可靠性分析，它可以围绕一个或一些特定的故障状态，作层层深入的分析，因而在清晰的失效树图形下，表达了系统的内在联系，并且明确表示部件与系统故障之间的逻辑关系。

2) 故障树分析法能把系统故障的各种因素联系起来。因此，有利于提高系统的可靠性，找出系统的薄弱环节。

3) 故障树建立后，对不曾参与系统设计的管理人员和维修人员来说相当于一个形象的管理、维修指南。因此，对培训长期使用的大系统人员更有意义。

4) 通过故障树可以定量地求出复杂系统的故障概率和其他可靠性参数。为改善和评估系统的可靠性提供有益的数据。

5) 故障树分析的发展是和计算机技术及其软件的发展紧密相关，图像信息技术也已经应用在故障树分

析之中，因此，编制计算机程序是故障分析中不可缺少的一部分。

6) 应用故障树分析可以求得复杂系统的最小路集和最小割集，便于计算系统的可靠性指标。

7) 故障树分析的理论基础是概率论、布尔代数和数理统计以及可靠性数学等数学工具。故障树分析法不仅应用于解决工程技术问题，而且也开始用于社会经济管理的系统工程分析。

故障树分析法还有些不足之处：

1) 建树复杂，需要具有丰富经验的工程技术人员参加。而且不同人所建造的“树”不会完全相同，有时可能发生错误。

2) 工作量大，耗费人力、物力较多，对于复杂系统的建树还存在一定的困难，有时建树时间过长。

所谓故障树分析法是把最不希望发生的故障状态作为故障分析的目标(在故障树中定义为顶事件)，因而导致这一故障状态所有可能发生的直接因素(在故障树中定义为中间事件)，再跟踪追溯找故障事件所有可能发生的直接原因，层层跟踪，直追寻到引起部件发生故障的全部原因(在故障树中定义为底事件)。用相应的代表符号及逻辑门把顶事件、中间事件、底事件联结成树形图，称此树形图为故障树。由上面分析可见，故障树就是事件之间的一个逻辑关系图，它表示了由元件、部件或人为失误等等事件，引起顶端事件的逻辑关系图。建立“树”以后，以故障树对系统的故障进行定性分析和定量计算，以故障树为工具对系统的失效进行评价的方法，称为故障树分析法。

### 5.2 基本符号

故障树中使用的符号分为事件符号与逻辑符号两类。

#### 5.2.1 事件符号

1) 矩形符号(图 45.4-19a)是故障树分析中的结果事件。可以是顶事件，也可以是由其他事件或事件组合所导致的中间事件。矩形事件的下端与逻辑门相联接，表示该事件是此逻辑门的一个输入。

2) 圆形符号(图 45.4-19b)是故障树分析中的基本事件，是分析中无需探明其发生原因的事件。

3) 菱形符号(图 45.4-19c)是故障树分析中的未探明事件，即原则上应进一步探明其原因但暂时不必或暂时不能探明其原因的事件。它又代表省略事件，一般地表示那些可能发生，但概率值微小的事件；或者对此系统到此为止不需要再进一步分析的故障事件，这些故障事件在定性分析中或定量计算中一般都可以忽略不计。

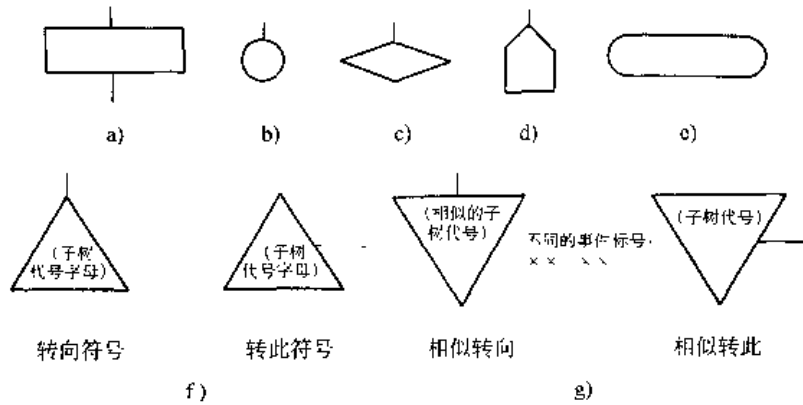


图 45.4-19 事件符号

4) 房形符号(图 45.4-19d)是开关事件,在正常工作条件下必然发生或必然不发生的事件,当房形中所给定的条件满足时,房形所在门的其他输入保留,否则除去。根据建树要求,可以是正常事件,也可以是故障事件。

5) 扁圆形符号(图 45.4-19e)是条件事件,是描述逻辑门起作用的具体限制的事件。

6) 三角形符号(图 45.4-19f、g)是转移符号,在故障树中经常出现条件完全相同,或是同一个故障事件在不同位置上出现,为了减少重复工作量及树的简化,用转移符号加上相应标号表示从某处转入,用转移符号加上相应标号表示转到某处。图 45.4-19f 为相同转移符号用以指明子树的位置,转向和转此字母代号相同,图 45.4-19g 相似转移符号用以指明相似于树的位置,转向和转此字母代号相同,事件的标号不同。

输出事件才发生。

6) 逻辑异或门(图 45.4-20f),逻辑异或门简称异或门,表示仅当单个输入事件发生时,输出事件才发生。

7) 逻辑禁门(图 45.4-20g),逻辑禁门简称禁门,它表示只有满足一定条件时,输入事件发生才能使输出事件发生的逻辑关系。一般用于表示某些非正常工作条件下发生的故障。

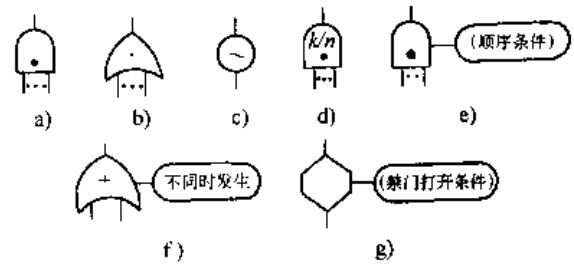


图 45.4-20 逻辑门符号

### 5.2.2 逻辑门符号

逻辑门符号与逻辑电路中所用符号相同,在故障树中只有两种状态,即成功与故障(用 1, 0 表示)。因此,一般地只用到几种基本的逻辑门。

1) 逻辑与门(图 45.4-20a),逻辑与门简称与门。它表示全部输入事件都发生才能使输出事件发生的逻辑关系。

2) 逻辑或门(图 45.4-20b),逻辑或门简称或门,它表示输入事件中只要有一个或多个发生就能使输出事件发生的逻辑关系。

3) 逻辑非门(图 45.4-20c),逻辑非门简称非门,它表示输出事件是输入事件的对立事件。

4) 逻辑表决门(图 45.4-20d),逻辑表决门简称表决门,它表示仅当  $n$  个输入事件中有  $k$  个或  $k$  个以上的事件发生时,输出事件才发生。

5) 逻辑顺序与门(图 45.4-20e),逻辑顺序与门简称顺序与门,表示仅当输入事件按规定的顺序发生时,

### 5.3 故障树的建立

故障树是故障分析(定性和定量分析)的对象,其完善程度直接影响分析结果的准确性,其简化的程度直接影响分析的工作量,也影响故障树的直观性。因此,建造故障树是故障树分析的关键。

对于大型复杂系统,由于故障机理交错多变,逻辑关系复杂,因此要求建树者必须慎重,必须广泛地掌握关于系统的设计、运行、安全分析等方面的知识和经验。所以,一般应由设计、工艺、试验、管理和使用等方面的专家组成一个精干的班子来完成建树任务。

目前建树的方法有两大类。一是人工建树——演绎法,一是计算机辅助建树。由于人工建树可以使建树者透彻地掌握事件的逻辑关系,这里只介绍它。

演绎法是从顶事件开始向下逐级分析的建树法。第一步,把顶事件排在第一行,用矩形符号表示出来;第二步,将导致顶事件发生的直接原因(包括硬件、软

件、环境因素和人为差错)并排在第二行,各用适当的事件符号表示出来(如底事件用圆形符号,省略事件用菱形符号,中间事件用矩形符号,……),并用适当的逻辑门同顶事件连接起来;第三步,分别分析上述中间事件发生的原因,将各自的原因并列在各自的下面作为第三行,用适当的事件符号表示之,并用适当的逻辑门同对应的上一级中间事件连接起来;……,直到把最基本的原因(底事件)都分析出来为止。

### 5.3.1 注意事项

为了使故障树能正确地、简明地把系统的故障事件表示出来,容易直观地看出系统的故障,因此在建造故障树时必须注意以下几点:

#### (1) 有明确的主流程, 确保逻辑严谨

所谓主流程是指能贯穿于系统各部件的功能故障,以其为纲,从顶事件到底事件逐渐分解建树。这样就可使建造故障树时思路明确,使人一目了然。

**例** 图 45.4-21 是一直流电动机驱动系统。其中:  $E$ ——24V 直流电源;  $S_1$ ——手动开关;  $S_2$ ——电磁开关;  $D$ ——电动机,额定电压  $U=24V$ ; 故障概率  $Q_D=0.001$ ;  $M$ ——液压泵,故障概率  $Q_M=0$ 。系统故障是不供油。

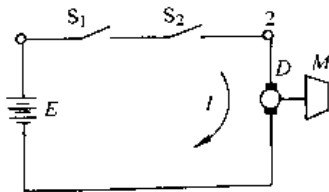


图 45.4-21 直流电动机驱动系统

以电动机不能启动为顶事件建造故障树,确定主流程。电动机不能启动可能是液压泵  $M$  卡住、电动机转子卡住、 $K_1$  和  $K_2$  未合上、电源故障和未给电动机额定电流。但是前面几种事件对此故障事件是独立事件不能作为主流程,只有电流是贯穿回路的。因为题中未告诉额定电流是多少;而额定电流是由额定电压决定的,因之最好以额定电压为主流程。

#### (2) 合理地确定边界条件,以便确定故障树的范围

所谓边界条件是指在建树前对系统、部件等提出的假设条件。通常有两类边界条件:

1) 系统的边界条件。它包括初始条件、已知的技术状态、已发生或正在发生的故障事件(含顶事件)、不允许出现的事件等。其中顶事件是最重要的边界条件之一。

2) 部件的边界条件。它包括假定部件所处的状态、

部件发生故障的概率等。如:

a) 确定不可能事件。一般把小概率事件当作不可能事件,建树时就不画出来了。

b) 确定必然事件。它是系统工作时在一定条件下必然发生的事件,或必然不发生的事件。

c) 确定某些事件发生的概率。

上述这些假定就是一组边界条件。有了边界条件,就明确了建树的范围。根据这些边界条件,在图 45.4-21 的例子中,就可选“电动机不能启动”为顶事件;否则就要选“油泵不供油”为顶事件了,但是,在确定边界条件时必须特别注意:

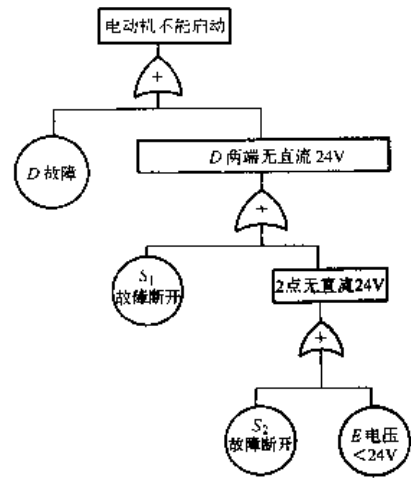


图 45.4-22 图 45.4-21 的故障树

a) 忽略小概率事件不等于忽略小部件的故障和小故障事件。这是两个不同的概念。

b) 有的事件概率虽小,但一旦发生后果严重,这种事件就不能作为不可能事件处理。

c) 精确地定义故障事件,必须做到只有一种解释,切忌多意性、模棱两可和含糊不清。否则,可能导致树中出现逻辑混乱、矛盾和错误。

d) 建树的头几步应考虑主要的,发生概率大的,或关键性的事件,随着分析的进展,再考虑次要的事件、发生概率较小的事件。

e) 严谨的逻辑性。系统中各故障事件的逻辑关系、条件必须分析清楚,不能紊乱和自相矛盾。

图 45.4 21 的故障树见图 45.4-22。

### 5.3.2 故障树简化

故障树简化的手段是去掉明显的逻辑多余事件和明显的逻辑多余门。往往凭直观即可进行简化,若利用布尔代数进行简化则更为合适。图 45.4-23 所示简化的例子下面即布尔代数的表达式。简化的另一手段是用相同的转移符号表示相同的子树,用相似的转移符号表示相似的子树。如图 45.4-24 所示。

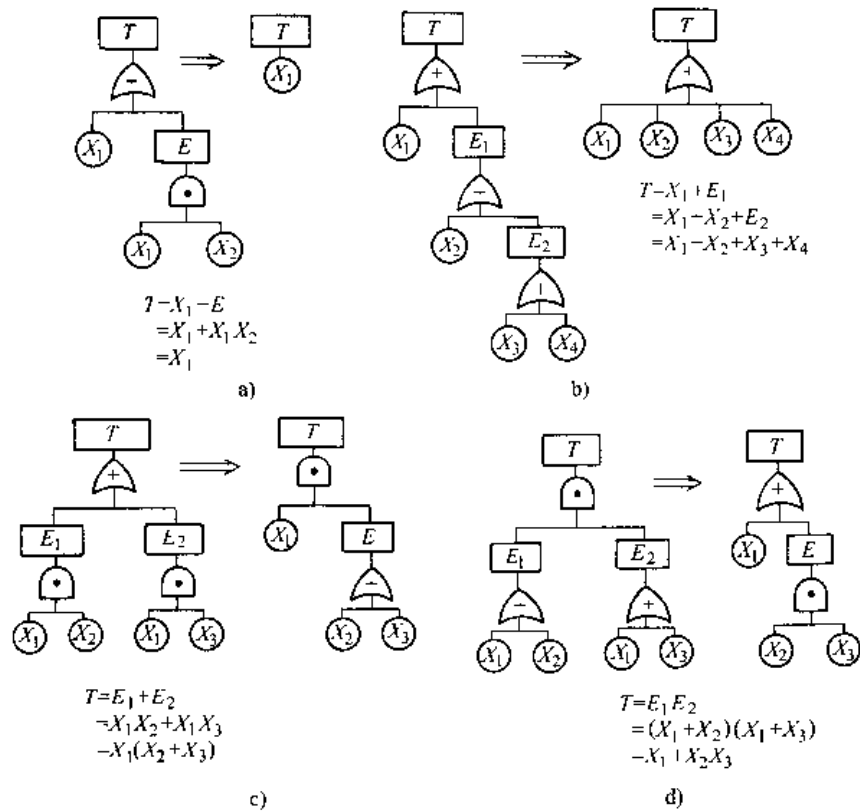


图 45.4-23 故障树的简化例

加建树工作。

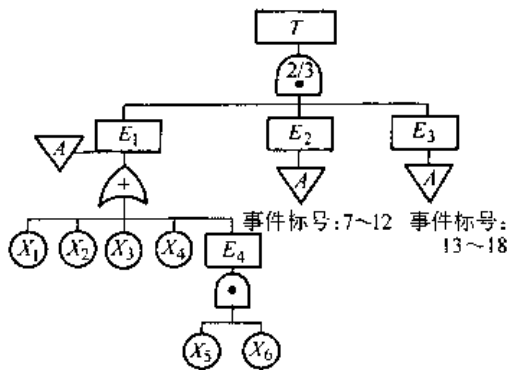


图 45.4-24 相似转移简化例

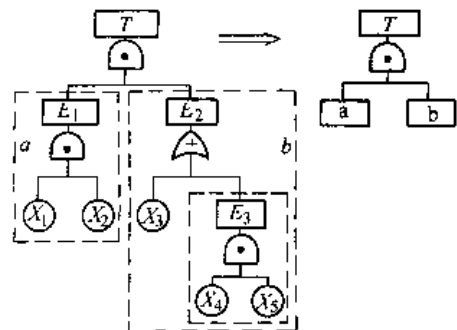


图 45.4-25 故障树的模块例

故障树的模块是故障树中若干个底事件的集合，各集合都不包含其他集合的底事件，即各集合中的底事件向上可到达同一个逻辑门，并且必须通过此门才能到达顶事件，故障树的其他底事件向上均不能到达该逻辑门，最大模块是没有其他模块包含它的模块。图 45.4-25 中的 a 和 b 均为最大模块。故障树的模块分解是找出故障树中尽可能大的模块，每个模块均构成一个模块子树，可单独进行定性和定量分析，并可用一等效的虚设底事件来代替，使原故障树的规模减小。

建树和简化不宜完全分为两个步骤，应该是边建树边简化，最后再全盘进一步简化。

建立故障树需对系统有深入的了解，因此必要时应聘请运行、维修、制造等各方面有经验的技术人员参

### 5.4 故障树的定性分析

故障树的定性分析主要是通过找出故障树中所有导致顶事件发生的最小割集。割集是导致正规故障树顶事件发生的若干底事件的集合。如割集中的任一底事件不发生，顶事件也就不发生，则这样的割集就是最小割集。例如图 45.4-26 中的故障树，若将各底事件分别用  $X_1, X_2, X_3, X_4$  表示，其割集有  $\{X_1\}, \{X_2\}, \{X_3, X_1\}, \{X_1, X_2\}, \{X_1, X_2, X_3\}, \dots, \{X_1, X_2, X_3, X_4\}$  等。显然，最小割集只是  $\{X_1\}, \{X_2\}$  和  $\{X_3, X_4\}$  三个。这也正表明能导致该系统故障的三种可能形式。组成最小割集的底事件个数称为最小割集

的阶。前面的最小割集中  $\{X_1\}$  和  $\{X_2\}$  都是一阶割集,  $\{X_3, X_4\}$  为二阶割集。一般阶数越低, 越容易发生故障, 因此最低阶的最小割集常是系统的薄弱环节。通过故障树的定性分析可洞察全系统的故障可能形式, 也可以发现系统的薄弱环节。

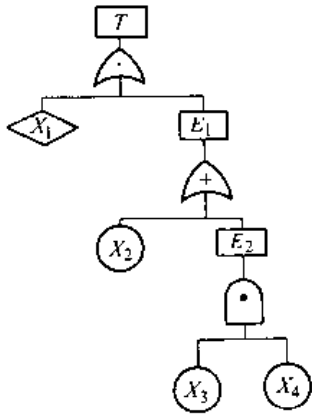


图 45.4-26 故障树的定性分析例

简单的故障树凭直观即可找出所有的最小割集。一般常用下行法或上行法求故障树的所有最小割集。

5.4.1 下行法

下行法的基本原则是: 对每一个输出事件, 若下面是或门, 则将该或门下的每一个输入事件各自排成一行; 若下面是与门, 则将该与门下的所有输入事件排在同一行。下行法的步骤是: 从顶事件开始, 由上向下逐级进行, 对每个结果事件重复上述原则, 直到所有结果事件均被处理, 所得每一行的底事件的集合均为故障树的一个割集。最后按最小割集的定义, 对各行的割集通过两两比较, 划去那些非最小割集的行, 剩下的即为故障树的所有最小割集。

例 对图 45.4-27 所示的故障树, 用下行法求其所有最小割集。

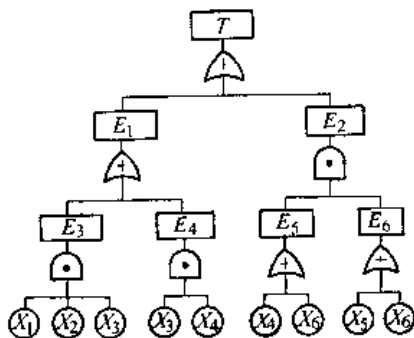


图 45.4-27 故障树的分析例

解 由顶事件  $T$  开始, 步骤如下, 见表 45.4-13。

1) 顶事件  $T$  下面是或门, 将该门下的输入事件  $E_1$  和  $E_2$  各自排成一行。

表 45.4-13 图 45.4-27 故障树下行法求最小割集

步 骤					
0	1	2	3	4	5
$T$	$E_1$ $E_2$	$E_3$ $E_4$ $E_7E_6$	$X_1X_2X_3$ $X_3X_4$ $X_1E_6$ $X_6E_6$	$X_1X_2X_3$ $X_3X_4$ $X_4X_5$ $X_4X_6$ $X_6X_5$ $X_5X_6=X_6$	$X_1X_2X_3$ $X_3X_4$ $X_4X_5$ $X_6$

2) 事件  $E_1$  下面是或门, 将该门下的输入事件  $E_3$  和  $E_4$  各自排成一行, 事件  $E_2$  下面是与门, 将该门下的输入事件  $E_5$  和  $E_6$  排在同一行。

3) 事件  $E_3$  下面是与门, 将该门下的输入事件  $X_1, X_2$  和  $X_3$  排在同一行; 事件  $E_4$  下面是与门, 将该门下的输入事件  $X_3$  和  $X_4$  排在同一行, 事件  $E_5$  下面是或门, 将该门下的输入事件  $X_4$  和  $X_6$  各自排成一行, 并与事件  $E_6$  组合成  $X_4E_6$  和  $X_6E_6$ 。

4) 事件  $E_6$  下面是或门, 将该门下的输入事件  $X_5$  和  $X_6$  各自排成一行, 并与事件  $X_4$  组合成  $X_4X_5$  和  $X_4X_6$ ; 与事件  $X_6$  组合成  $X_5X_6$  和  $X_6X_6$ 。至此, 故障树的所有结果事件都已被处理, 这时所得的每行均为一个割集。

5) 进行两两比较, 因为  $\{X_6\}$  是割集, 故  $\{X_4, X_5\}$  和  $\{X_3, X_6\}$  不是最小割集, 应划去。最后得该故障树的所有最小割集为:

$$\{X_6\}, \{X_3, X_4\}, \{X_4, X_5\}, \{X_1, X_2, X_3\}$$

5.4.2 上行法

上行法的基本原则是: 对每个结果事件, 若下面是或门, 则将此结果事件表示为该或门下的各输入事件的布尔和 (事件并); 若下面是与门, 则将此结果事件表示为该与门下的输入事件的布尔积 (事件交)。上行法的步骤是: 从底事件开始由下向上逐级进行, 对每个结果事件重复上述原则, 直到所有结果事件均被处理。将所得的表达式逐次代入, 按布尔运算的规则, 将顶事件表示成底事件积之和的最简式, 其中每一项对应于故障树的一个最小割集, 从而求得故障树的所有最小割集。

例 对图 45.4-27 所示的故障树, 用上行法求其所有最小割集。

解 从底事件开始,

$$E_3 = X_1X_2X_3$$

$$E_4 = X_3X_4$$

$$E_5 = X_4 + X_6$$

$$E_6 = X_5 + X_6$$

$$E_1 = X_3 + X_4 = X_1X_2X_3 + X_3X_4$$

$$E_2 = X_5X_6 = (X_4 + X_6)(X_5 + X_6)$$

$$= X_3X_2 + X_1X_6 + X_7X_8 + X_6X_9$$

$$= X_4X_5 + X_6$$

$$T = E_1 + E_2 = X_1X_2X_3 + X_4X_5 + X_6X_7 + X_6$$

故得故障树的所有最小割集为

$$\{X_1, X_2, X_3\}, \{X_4, X_5\}, \{X_6, X_7\}, \{X_6\}$$

### 5.5 故障树的定量分析

如有足够的数据能够估计出故障树中各底事件发生的概率,则在所有底事件相互独立的条件下,可对故障树进行下述定量分析。

#### 5.5.1 顶事件发生的概率

求顶事件发生的概率有多种方法,现仅介绍下面两种。

##### (1) 直接概率法

本法根据故障树的或门相当于可靠性框图的串联模型,与门相当于可靠性框图的并联模型。如图 45.4-28a 所示的或门

$$P(E) = q_E = 1 - \prod_{i=1}^n (1 - q_i) \quad (45.4-43)$$

图 45.4-28b 所示的与门

$$P(E) = q_E = \prod_{i=1}^n q_i \quad (45.4-44)$$

式中  $q_i$ ——故障树中底事件  $X_i$  发生的概率  $P(X_i)$ 。

一般由下而上求各层中间事件在内的各事件发生的概率直至求得顶事件发生的概率。应该注意,用本方法时不仅要求所有底事件相互独立,而且同一底事件在故障树中只能出现一次。当同一底事件在故障树中出现不止一次,则应先按图 45.4-23 的原则进行简化。

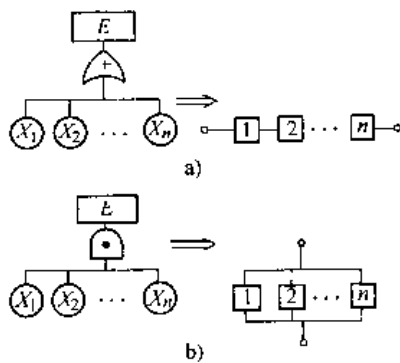


图 45.4-28 与可靠性框图相应的或门及与门

例 图 45.4-29a 所示的故障树,已知各底事件发生的概率  $q_1 = q_2 = q_3 = 0.1$ , 求其顶事件发生的概率。

解 因基本事件  $X_1$  在故障树中不止一次出现,故按图 45.4-23 简化为图 45.4-29b, 再按简化后的故障树由下而上计算于下

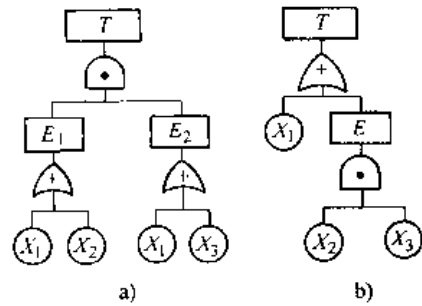


图 45.4-29 图 45.4-23 的故障树

$$q_T = q_2q_3 = 0.1 \times 0.1 = 0.01$$

$$P(T) = Q = 1 - (1 - q_1)(1 - q_T) = 1 - (1 - 0.1)(1 - 0.01) = 0.109$$

##### (2) 不交布尔代数法

本法可用于故障树的规模比较大的情况,而且不要求底事件在故障树中只出现一次。用本方法求顶事件发生概率的步骤如下:

- 1) 用上行法求最小割集,并将顶事件表示为各底事件积之和的最简布尔表达式。
- 2) 将所得最简布尔表达式化为互不相交的布尔和。
- 3) 对求得的不交化表达式两边求概率,即得顶事件发生的概率。

例 2 求图 45.4-27 故障树顶事件发生的概率。已知各底事件发生的概率如下

$q_1$	$q_2$	$q_3$	$q_4$	$q_5$	$q_6$
0.02	0.02	0.03	0.025	0.025	0.01

解 顶事件表示为各底事件积之和的最简布尔表达式为

将上述化为互不相干的布尔和为

$$T = X_6 + X_3X_4 + X_4X_5 + X_1X_2X_3$$

将上述化为互不相交的布尔和为

$$\begin{aligned} T &= X_6 + X_3X_4\bar{X}_6 + X_4X_5\bar{X}_6\bar{X}_3 + \\ &X_1X_2X_3\bar{X}_6\bar{X}_4(\bar{X}_4 + X_1\bar{X}_5) \\ &= X_6 + X_3X_4\bar{X}_6 + X_4X_5\bar{X}_6\bar{X}_3 \\ &+ X_1X_2X_3\bar{X}_6\bar{X}_4 \end{aligned}$$

其中,  $\bar{X}_1$  表示底事件  $X_1$  的对立事件,即表示第  $i$  个底事件不发生。

将上面不交化表达式两边求概率,得顶事件发生的概率为

$$\begin{aligned} P(T) = Q &= P(X_6) + P(X_3X_4\bar{X}_6) + \\ &P(X_4X_5\bar{X}_6\bar{X}_3) + P(X_1X_2X_3\bar{X}_6\bar{X}_4) \\ &= q_6 + q_3q_4(1 - q_6) + q_4q_5(1 - q_6)(1 - q_3) + \\ &q_1q_2q_3(1 - q_6)(1 - q_4) \\ &= 0.01 + 0.03 \times 0.025(1 - 0.01) + 0.025 \times \\ &0.025(1 - 0.01)(1 - 0.03) + 0.02 \times \end{aligned}$$



$$0.02 \times 0.03(1 - 0.01)(1 - 0.025) \\ \dots 0.011354$$

### 5.5.2 重要度

重要度是评价故障树中各底事件对系统顶事件发生影响大小的尺度。按不同定义有多种重要度, 现仅介绍下述三种。

#### (1) 底事件的概率重要度

第  $i$  个底事件的概率重要度表示, 当第  $i$  个底事件发生概率的微小变化而导致顶事件发生概率的变化率。第  $i$  个底事件的概率重要度为

$$I_P(i) = \frac{\partial Q}{\partial q_i} \quad i = 1, 2, \dots, n \quad (45.4-45)$$

式中  $Q$ ——顶事件发生的概率, 它是各底事件发生概率  $q_1, q_2, \dots, q_n$  的函数  $Q(q_1, q_2, \dots, q_n)$ 。

#### (2) 底事件的相对概率重要度

第  $i$  个底事件的相对概率重要度表示, 当第  $i$  个底事件发生概率微小的相对变化  $\left\{ \frac{\Delta q_i}{q_i} \right\}$  而导致顶事件发生概率的相对变化  $\left( \frac{\Delta Q}{Q} \right)$  率。第  $i$  个底事件的相对概率重要度为

$$I_r(i) = \frac{q_i}{Q} \frac{\partial Q}{\partial q_i} = \frac{q_i I_P(i)}{Q} \quad (45.4-46)$$

式中  $I_P(i)$ ——第  $i$  个底事件的概率重要度。

#### (3) 底事件的结构重要度

第  $i$  个底事件的结构重要度表示, 当第  $i$  个底事件由正常变为故障引起顶事件由正常变为故障状态数与该底事件正常状态数之比。底事件的结构重要度从故障树结构的角度反映了各底事件在故障树中的重要程度。第  $i$  个底事件的结构重要度为

$$I_s(i) = \frac{n_s(i)}{2^{n-1}} \quad (45.4-47)$$

式中  $n$ ——故障树中底事件总数;

$n_s(i)$ ——第  $i$  个底事件由正常变为故障导致顶事件由正常变为故障的状态数。

为理解式 (45.4-46), 以图 45.4-29 的故障树为例来说明。该故障树有 3 个底事件, 可绘得底事件和顶事件的全部状态如表 45.4-14 所示。由表可看出: 底事件总数  $n=3$ , 各个底事件正常状态数均为 4, 即式 (45.4-44) 中  $2^{n-1}=4$ ; 第 1 个底事件由正常变为故障导致顶事件由正常变为故障状态数有  $(^0_100)$ ,  $(^0_101)$ ,  $(^0_110)$  三个, 故  $n_s(1)=3$ ; 类似, 第 2 个底事件有  $(0^1_21)$  一个, 故  $n_s(2)=1$ ; 第 3 个底事件有  $(01^3_3)$  一个, 故  $n_s(3)=1$ 。代入式 (45.4-44) 得

$$I_P(1) = \frac{3}{4}, I_P(2) = \frac{1}{4}, I_P(3) = \frac{1}{4}$$

表 45.4-14 状态真值表

(0—正常, 1—故障)			顶事件 $T$
底事件			
$X_1$	$X_2$	$X_3$	
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

用真值表求出  $I_P(i)$  门时如底事件较多就很不方便。理论上已经证明, 当所有底事件发生的概率都取为  $1/2$  时, 则底事件的概率重要度等于底事件的结构重要度。故可将  $q_i=1/2$  代入式 (45.4-44) 即可求得结构重要度。

例 求图 45.4-27 故障树各底事件的概率重要度、相对概率重要度和结构重要度。

解 该故障树顶事件发生概率为

$$Q = q_6 - q_1 q_4 (1 - q_6) + q_4 q_5 (1 - q_6) (1 - q_3) + \\ q_1 q_2 q_3 (1 - q_4) (1 - q_6) = 0.011354$$

a) 底事件概率重要度, 将上式代入式 (45.4-44) 得

$$I_P(1) = q_2 q_4 (1 - q_5) (1 - q_6) \\ = 0.02 \times 0.03 (1 - 0.025) (1 - 0.01) \\ = 0.0005791$$

$$I_P(2) = q_1 q_3 (1 - q_4) (1 - q_6) \\ = 0.02 \times 0.03 (1 - 0.025) (1 - 0.01) \\ = 0.0005791$$

$$I_P(3) = q_4 (1 - q_5) - q_1 q_2 (1 - q_6) + \\ q_1 q_2 (1 - q_4) (1 - q_6) \\ = 0.025 (1 - 0.01) - \\ 0.025 \times 0.025 (1 - 0.01) + \\ 0.02 \times 0.02 \times \\ (1 - 0.025) (1 - 0.01) = 0.02452$$

$$I_P(4) = q_3 (1 - q_6) + q_5 (1 - q_6) (1 - q_3) - \\ q_1 q_2 q_3 (1 - q_6) \\ = 0.03 (1 - 0.01) + \\ 0.025 (1 - 0.01) (1 - 0.03) - \\ 0.02 \times 0.02 \times 0.03 (1 - 0.01) \\ = 0.05370$$

$$I_P(5) = q_1 (1 - q_6) (1 - q_3) \\ = 0.025 (1 - 0.01) (1 - 0.03) \\ = 0.02101$$

$$I_P(6) = 1 - q_1 q_4 - q_4 q_5 (1 - q_3) - q_1 q_2 q_3 (1 - q_4) \\ = 1 - 0.03 \times 0.025 - 0.025 \times 0.025 \times$$

$$(1 - 0.03) - 0.02 \times \\ 0.02 \times 0.03(1 - 0.025) \\ = 0.99863$$

b) 底事件的相对概率重要度, 按式 (45.4-15) 得

$$I_C(1) = \frac{q_1 I_p(1)}{Q} = \frac{0.02 \times 0.0005791}{0.011354} \\ = 0.001020$$

$$I_C(2) = \frac{q_2 I_p(2)}{Q} = \frac{0.02 \times 0.0005791}{0.011354} \\ = 0.001020$$

$$I_C(3) = \frac{q_3 I_p(3)}{Q} = \frac{0.03 \times 0.02452}{0.011354} \\ = 0.06479$$

$$I_C(4) = \frac{q_4 I_p(4)}{Q} = \frac{0.025 \times 0.05370}{0.011354} \\ = 0.1182$$

$$I_C(5) = \frac{q_5 I_p(5)}{Q} = \frac{0.025 \times 0.02401}{0.011354} \\ = 0.05287$$

c) 底事件的结构重要度, 取所有底事件发生的概率均为  $q_i = 1/2$  代入式 (45.4-44), 见前面已求得各底事件概率重要度, 故可得

$$I_s(1) = q^1 = \frac{1}{2^1} = \frac{1}{16}$$

$$I_s(2) = q^1 = \frac{1}{2^1} = \frac{1}{16}$$

$$I_s(3) = q^2 - q^1 + q^1 = \frac{1}{2^2} - \frac{1}{2^3} + \frac{1}{2^4} = \frac{3}{16}$$

$$I_s(4) = q^2 + q^1 - q^1 = \frac{1}{2^2} + \frac{1}{2^3} = \frac{5}{16}$$

$$I_s(5) = q^1 = \frac{1}{2^3} = \frac{1}{8}$$

$$I_s(6) = 1 - q^2 - q^3 - q^4 = 1 - \frac{1}{2^2} - \frac{1}{2^3} - \frac{1}{2^4} \\ = \frac{9}{16}$$

由重要度的计算可明显看出底事件  $X_6$  是影响该系统可靠性最重要的环节, 其他底事件重要程度则按其重要度的数值即可得到反映。

## 第 5 章 维修性设计

### 1 概述

工程实际中的系统多数属于可以修复的系统。可靠性考虑的是使设计出来的产品经久耐用不发生故障的问题；而就维修性来说，关心的重点则是产品在发生故障时能够使故障尽快排除，或者减少和预防故障的发生。将可靠性与维修性结合起来，称为可修复系统（产品）的广义可靠性，又称作有效性或可用性。

维修性的定义是：“在规定条件下使用的产品在规定的时间内，按规定的程序和方法进行维修时，保持或恢复到能完成规定功能的能力。”如果用概率来表示和量度这种能力，就是维修度，用  $M(t)$  表示。维修性参数与可靠性参数一样，也是随机变量、概率参数，用数理统计理论和概率论进行分析；但可修复系统的两种状态（正常状态与故障状态）之间的相互转移情况，比不可修复系统要复杂，需借助随机过程理论来描述，对产品维修性的研究，是可靠性研究的进一步发展。良好的维修性不仅是改善、维持或提高系统功能的主要技术措施之一，而且使产品在使用中发挥最大效用。因此，维修系统的可靠性技术，既是产品设计的重要内容之一，也是设备管理和设备运行的主要内容。

#### 1.1 维修性的特征量

1) 维修度  $M(t)$  维修度是维修时间的函数， $M(t)$  称为维修度函数。

如果用随机变量  $T$  表示产品从开始维修到修复的时间，其概率密度为  $m(t)$ ，则

$$M(t) = P(T > t) = \int_0^{\infty} m(t) dt \quad (45.5-1)$$

2) 修复率 修复率是修理时间已达到某个时刻尚未修复的产品，在该时刻后的单位时间内完成修理的概率，记为  $\mu(t)$

$$\mu(t) = \frac{m(t)}{1 - M(t)} \quad (45.5-2)$$

3) 平均修复时间(MTTR) 修复时间是“从发现故障到产品恢复规定功能所需的时间，即故障诊断、修理准备及修理实施时间之和”。

平均修复时间为修复时间的均值，记为  $\bar{t}$  或 MTTR

$$\bar{t} = \int_0^{\infty} t m(t) dt \quad (45.5-3)$$

4) 有效度  $A(t)$  瞬时有效度是“产品在某时刻具

有或维持其规定功能的概率”。当时间趋于无限时，瞬时有有效度的极限值称为稳态有效度。在某个观察时期内，产品能工作的时间对能工作时间与不能工作时间之和的比，为有效度的观测值，表示为

$$A(t) = \frac{U}{U + D} \quad (45.5-4)$$

式中  $U$ ——产品能工作的时间；

$D$ ——产品不能工作的时间。

在某个规定的时间区间内，有效度的平均值称为平均有效度。

有效性是可靠性与维修性的综合，因而是广义的可靠性。当产品的失效寿命为指数分布时，由式(45.5-4)可得

$$A(t) = \frac{MTBF}{MTBF + MTTR} = \frac{\mu}{\mu + \lambda} \quad (45.5-5)$$

式中 MTBF——平均无故障工作时间；

MTTR——平均修复时间；

$\mu$ ——修复率；

$\lambda$ ——失效率。

由式(45.5-5)知，若要使  $A(t)$  增大，必须提高 MTBF 值和减小 MTTR 值。

5) 重要度 重要度是系统中某设备发生的故障次数与整个系统中所有设备发生的故障次数之比

$$\text{重要度} = \frac{\text{某设备故障引起的系统故障次数}}{\text{系统发生故障的次数}} \quad (45.5-6)$$

6) 经济尺度 常用的有

$$\text{费用比} = \frac{\text{全年维修费}}{\text{购置费}} \quad (45.5-7)$$

$$\text{单位产品的维修费用} = \frac{\text{维修费用}}{\text{产品产量}} \quad (45.5-8)$$

7) 量大维修时间  $t_{M_{max}}$  与维修度  $M(t) = 0.95$  相应的维修时间，意即在规定的最大维修时间  $t_{M_{max}}$  内完成维修措施的概率为 95%，或所有维修措施的 95% 将能在  $t_{M_{max}}$  内完成，如图 45.5-1 所示。由图 45.5-1 知， $t_{M_{max}}$  即  $t_{0.95}$ 。

当修复时间服从指数分布时， $t_{M_{max}} = 2.9957$  MTTR。如果 MTTR = 1h，则  $t_{M_{max}} \approx 3h$ ，如 MTTR = 0.5h， $t_{M_{max}} = 1.5h$ ，如图 45.5-2 所示。

8) 设备修复时间 设备修复时间分布的中位值，即与  $M(t) = 0.50$  相应的维修时间，如图 45.5-1 所示，ERT 即  $t_{0.5}$ 。在 ERT 内，所有维修措施的 50% 将能完

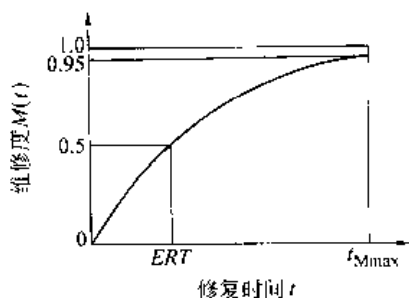


图 45.5-1  $M(t)$   $t$  的关系曲线

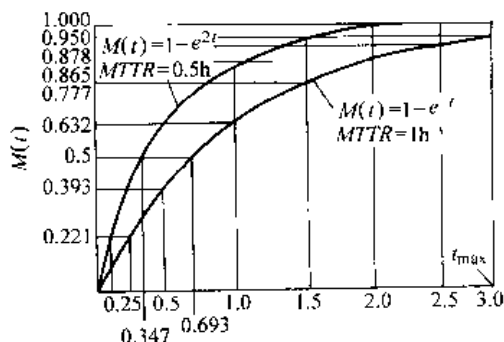


图 45.5-2 MTTR 与相应的  $t_{Mmax}$

成。

$$m = \frac{\sum_{i=1}^n \lambda_i \ln t_i}{\sum_{i=1}^n \lambda_i} \quad (45.5-9)$$

当修复时间分布为不同的概率分布类型时，和 ERT 与 MTTR 的关系式列于表 45.5-1。

表 45.5-1  $t_{Mmax}$  和 ERT 与 MTTR 的关系式

分布类型	$t_{Mmax}$	ERT
指数	3MTTR	0.7MTTR
正态	MTTR + 1.65 $\sigma_t$	MTTR
对数正态	anti ln( $m + 1.65\sigma_t$ )	$\frac{-\sigma_t^2}{e} \times MTTR$

注： $m$ ——修复时间  $t_i$  对数的均值。

## 1.2 维修的分类

维修可以分为事后维修和预防维修两大类。

### 1.2.1 事后维修

事后维修是未列入计划的维修。当元件、零件、设备或系统发生故障时，再进行维修，因为它常常发生在设备或系统正在正常运行而最不希望发生故障的时刻，所以一般不宜采用。只有在采用预防维修所得到的经济效益不足以支付增加的费用时，才采用事后维修。

### 1.2.2 预防维修

预防维修是为防止产品性能退化或降低产品故障

的概率，按事前规定的计划或相应技术条件的规定进行的维修。对于一旦发生故障就会造成重大经济损失甚至人身事故的场合，以及安全性要求很高的设备和系统，预防维修尤其必要。

由图 45.5-3 可知，预防维修可以改善产品耗损故障期内的效率，因而可以提高设备的可靠性。这样设备的可靠性由提供的预防维修来决定。因此，对可靠性的研究导致了维修性的研究。

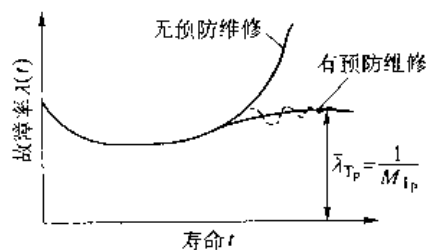


图 45.5-3 预防维修对故障率的影响

## 1.3 维修性设计与可靠性设计的关系

由式 (45.5-5) 可知，可靠性与维修性都是决定产品有效性的因素，但应当首先考虑提高产品的可靠性。

进行维修性设计时所依据的数学基础与可靠性设计是一样的。所以，在分析维修度函数时，只要将有关可靠度函数的几个变量进行适当代换，即可得到维修度函数。具体地说，这些互相对应的变量是：

- 1) 可靠性设计中的失效时间分布  $f(t)$  与维修修复时间分布  $m(t)$ 。
- 2) 可靠性设计中的失效率  $\lambda(t)$  和平均无故障工作时间 MTBF，与维修性设计中的修复率  $\mu(t)$  和平均修复时间 MTTR。
- 3) 可靠性设计中的失效概率（不可靠度） $F(t)$  与维修性设计中的维修度  $M(t)$ 。

维修性和可靠性特征量的对应关系如图 45.5-4 和表 45.5-2 所示。图中， $F(t)$  与  $M(t)$  相对应， $F(t)$  越高表示失效概率越高， $M(t)$  越高表示修复概率越高。失效与修复，其效果是对立的。就广义可靠性而言， $F(t)$  越低， $M(t)$  越高，则可靠性越佳。平均修复时间、平均修复率等观测值与对应的平均寿命、平均失效率等观测值计算法均类似。

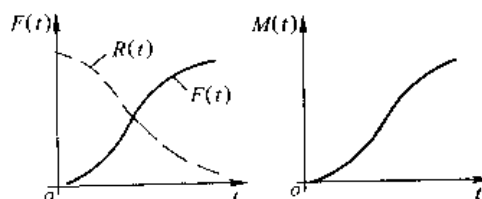


图 45.5-4 不可靠度与维修度函数

例 由 5 台同样设备维修记录知，共维修 35 次，

总计维修时间 168h,求平均修复时间和平均修复率观测值。若维修时间为指数分布,估计维修时间为 10h 的维修度和维修度为 0.95 的修复时间。

表 45.5-2 可靠性与维修性对应关系

项 目	可靠性	维修性	
累积分布函数	$F(t)=1-R(t)$ $-\int_0^t f(x)dx$	$M(t)=\int_0^t m(x)dx$	
概率密度	$f(t)=\frac{dF(t)}{dt}$	$m(t)=\frac{dM(t)}{dt}$	
失效率和修复率	$\lambda(t)=\frac{f(t)}{1-F(t)}$	$\mu(t)=\frac{m(t)}{1-M(t)}$	
指数分布	累积分布	$F(t)=1-e^{-\lambda t}$	$M(t)=1-e^{-\mu t}$
	平均时间	MTTF= $\frac{1}{\lambda}$ (MTBF)	MTTR= $\frac{1}{\mu}$

解 求平均修复时间的观测值

$$\hat{\tau} = \frac{1}{t} \cdot \sum t = \frac{168}{35} \text{h} = 4.8 \text{h}$$

求平均修复率的观测值

$$\hat{\mu} = \frac{r}{\sum t} = \frac{35}{168} \text{次/h} = 0.2083 \text{次/h}$$

若维修时间为指数分布,求维修时间为 10h 的维修度,按表 45.5-2 知

$$\hat{M}(t) = 1 - e^{-\hat{\mu}t} = 1 - e^{-0.2083 \times 10} = 0.875$$

若要求  $M(t) = 0.95$ , 则由

$$\hat{M}(t) = 1 - e^{-\hat{\mu}t} = 1 - e^{-0.2083t} = 0.95$$

$$\text{可解得 } \hat{\tau}_{0.95} = 14.38 \text{h}$$

通常将  $t_{0.95}$  称为最大修复时间,其实若规定  $t_{0.95}$  为修复时间,则超过该时间的维修次数仍有 5%。

## 2 修复时间分布和维修度函数

### 2.1 修复时间分布

维修时间是一个随机变量,一般认为维修时间的分布为对数正态分布。

例 已知在系统中的一个具体零件的故障和修复情况。其修复时间的数据如表 45.5-3 所示,试确定系统的修复时间分布,包括:①确定当零件失效时系统修复时间的概率密度函数,并绘图表示之;②确定系统的平均修复时间(MTTR);③确定系统修复时间分布的中位值;④完成修复工作的一半所需的时间;⑤确定修复时间分布的最大频值;⑥确定修复时间分布的标准差。

解 在大多数情况下修复时间呈对数正态分布,为了确定表 45.5-3 中给出的修复时间的概率密度函

数,首先

1) 计算修复时间的对数的均值  $\bar{t}'$  和标准差  $s'_t$

$$\bar{t}' = \overline{\ln t} = \frac{\sum_{j=1}^{N'} n_j t'_j}{N} = \frac{\sum_{j=1}^{N'} n_j t'_j}{\sum_{j=1}^{N'} n_j} \quad (45.5-10)$$

表 45.5-3 系统的修复时间数据

组号	修复时间 $t$ /h	观察频数 $n$	组号	修复时间 $t$ /h	观察频数 $n$
1	0.2	1	16	3.3	2
2	0.3	1	17	4.0	2
3	0.5	4	18	4.5	1
4	0.6	2	19	4.7	1
5	0.7	3	20	5.0	1
6	0.8	2	21	5.4	1
7	1.0	4	22	5.5	1
8	1.1	1	23	7.0	1
9	1.3	1	24	7.5	1
10	1.5	4	25	8.8	1
11	2.0	2	26	9.0	1
12	2.2	1	27	10.3	1
13	2.5	1	28	22.0	1
14	2.7	1	29	24.5	1
15	3.0	2			

式中  $n_j$ ——观察频数,如表 45.5-3 中第 3 栏所示;

$N'$ ——数据组的数目,对于本例  $N' = 29$ ;

$t_j$ ——修复时间,如表 45.5-3 中第 2 栏所示;

$t'_j$ ——修复时间  $t_j$  的对数,即  $t'_j = \ln t_j$ ;

$N$ ——总的观察频数,对于本例,  $N = \sum_{j=1}^{N'} n_j = 46$ 。

$$s'_t = \left[ \frac{\sum_{j=1}^{N'} n_j (t'_j)^2 - N(\bar{t}')^2}{N-1} \right]^{\frac{1}{2}} \quad (45.5-11)$$

计算数据列于表 45.5-4 中,将表 45.5-4 中的有关数据代入式(45.5-10、11),得

$$\bar{t}' = \frac{\sum_{j=1}^{N'} n_j t'_j}{N} = \frac{30.30439}{46} \text{h} = 0.65879 \text{h}$$

$$s'_t = \left[ \frac{75.84371 - 46 \times (0.65879)^2}{46-1} \right]^{\frac{1}{2}} \text{h} = 1.11435 \text{h}$$

修复时间分布的概率密度函数为

$$m(t) = \frac{1}{ts_s \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln t}{s'_t} \right)^2} = \frac{1}{t(1.11435) \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln t - 0.65879}{1.11435} \right)^2} \quad (45.5-12)$$

对于不同的修复时间,  $m(t)$  的值列于表 45.5-5, 根据表 45.5-5 中的值,可以绘出  $m(t)$  曲线,如图 45.5-5 所示。由图 45.5-5 可知,修复时间分布  $m(t)$  是向右倾斜的对数正态分布。

表 45.5-4 计算  $\bar{t}'$  和  $s_1$  所需的有关数据

组号	修复时间 $t$ (小时)	$t' = \ln t$	$(t')^2$	观察频数 $n$	$nt'$	$n(t')^2$
1	0.2	-1.60944	2.59029	1	1.60944	2.59029
2	0.3	1.20397	1.44955	1	-1.20397	1.44955
3	0.5	-0.69315	0.48045	4	-2.77260	1.92180
4	0.6	-0.51083	0.26094	2	-1.02166	0.52188
5	0.7	-0.35667	0.12721	3	1.07001	0.38166
6	0.8	-0.22314	0.04979	2	-0.44628	0.09958
7	1.0	0.00000	0.00000	4	0.00000	0.00000
8	1.1	0.09531	0.00908	1	0.09531	0.00908
9	1.3	0.26236	0.06884	1	0.26236	0.06884
10	1.5	0.40547	0.16440	4	1.62188	0.65760
11	2.0	0.69315	0.48045	2	1.38630	0.96090
12	2.2	0.78846	0.62167	1	0.78846	0.62167
13	2.5	0.91629	0.83959	1	0.91629	0.83959
14	2.7	0.99325	0.98655	1	0.99325	0.98655
15	3.0	1.09861	1.20693	2	2.19722	2.41390
16	3.3	1.19392	1.42545	2	2.38784	2.85090
17	4.0	1.38629	1.92181	2	2.77258	3.84362
18	4.5	1.50408	2.26225	1	1.50408	2.26225
19	4.7	1.54756	2.39495	1	1.54756	2.39495
20	5.0	1.60944	2.59029	1	1.60944	2.59029
21	5.4	1.68640	2.84394	1	1.68640	2.84394
22	5.5	1.70475	2.90617	1	1.70475	2.90617
23	7.0	1.94591	3.78657	1	1.94591	3.78657
24	7.3	2.01490	4.05983	1	2.01490	4.05983
25	8.8	2.17475	4.72955	1	2.17475	4.72955
26	9.0	2.19722	4.82780	1	2.19722	4.82780
27	10.3	2.33214	5.43890	1	2.33214	5.43890
28	22.0	3.09104	9.55454	1	3.09104	9.55454
29	54.5	3.19867	10.23151	1	3.19867	10.23151
$N' = 29$				$N = \sum_{j=1}^{N'} n_j$ = 46	$\sum_{j=1}^{N'} n_j t'_j$ = 30.30439	$\sum_{j=1}^{N'} n_j (t'_j)^2$ = 75.84371

表 45.5-5 修复时间的概率密度函数  $m(t)$  和  $m(t')$  (按表 45.5-3 数据算出)

修复时间 $t$	对数正态概率密度函数 $m(t)$	正态概率密度函数 $m(t') = m(\ln t)$	修复时间 $t$	对数正态概率密度函数 $m(t)$	正态概率密度函数 $m(t') = m(\ln t)$
0.02	0.00398	$7.95 \times 10^{-5}$	4.0	0.07232	0.28929
0.1	0.10480	0.01048	4.4	0.06195	0.27258
0.2	0.22552	0.04510	5.0	0.04976	0.24880
0.3	0.29510	0.08853	6.0	0.03559	0.21351
0.5	0.34300	0.17150	7.0	0.02625	0.18373
0.7	0.33770	0.23639	8.0	0.01985	0.15884
1.0	0.30060	0.30060	9.0	0.01534	0.13804
1.4	0.24524	0.34334	10.0	0.01206	0.12061
1.8	0.19849	0.35728	20.0	0.00199	0.03971
2.0	0.17892	0.35784	30.0	0.00058	0.01733
2.4	0.14638	0.35130	40.0	—	0.00888
3.0	0.11039	0.33118	80.0	—	0.00132
3.4	0.09260	0.31483			

同理,修复时间的对数  $t'$  的概率密度函数为

$$m(t') = m(\ln t) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t' - \bar{t}'}{\sigma_t} \right)^2} \quad (45.5-13)$$

$$m(t') = \frac{1}{1.11435 \times \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t' - 0.65829}{1.11435} \right)^2}$$

对于不同的修复时间,  $m(t')$  的值列于表 45.5-5。

根据表 45.5-5 中的值,可以绘出图 45.5-6 所示的

$m(t')$  曲线。由图 45.5.6 可知,  $m(t')$  为一正态分布曲线。

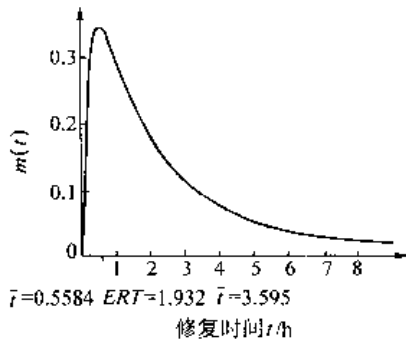


图 45.5.5 修复时间分布  $m(t)$

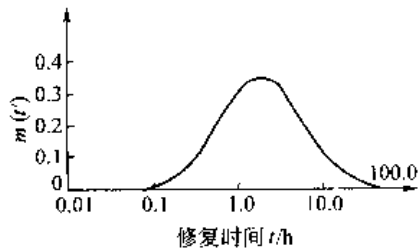


图 45.5.6 修复时间对数的分布  $m(t')$

2) 系统的平均修复时间(MTTR) $\bar{t}$ 由下式得出

$$\bar{t} = e^{\bar{t}' + \frac{1}{2}(s_t')^2} = e^{0.65479 + \frac{1}{2}(1.11475)^2} \text{h} = 3.595\text{h} \quad (45.5-14)$$

3) 系统修复时间的中位值为

$$ERT = e^{\bar{t}'} = e^{0.69879} = 1.932\text{h} \quad (45.5-15)$$

4) 完成修复工作的一半所需的时间,就是系统修复时间的中位值,表示为

$$t_{0.50} = ERT = 1.932\text{h}$$

5) 修复时间的最大频值是最经常发生的修复这一系统的时间,表示为

$$t = e^{\bar{t}' - s_t'^2} = e^{0.65879 - (1.11435)^2} = 0.358\text{h} \quad (45.5-16)$$

6) 修复时间的标准差为

$$s_t = \bar{t}(e^{s_t'^2} - 1) = 3.595(e^{1.11435^2} - 1)^{1/2} = 5.64\text{h} \quad (45.5-17)$$

## 2.2 维修度函数

### 2.2.1 已知修复时间为对数正态分布时的维修度函数

由维修度的定义可知,一旦修复时间分布的概率密度函数已经确定,相应于规定时间  $t_1$  的维修度为

$$M(t_1) = P(t \leq t_1) = \int_0^{t_1} m(t) dt \quad (45.5-18)$$

式中  $m(t)$  修复时间的概率密度函数。

$M(t_1)$  与  $m(t)$  的关系如图 45.5-7 所示。图中的

$G(t_1)$  为不可维修度。

$$G(t_1) = 1 - M(t_1) \quad (45.5-19)$$

当  $m(t)$  为对数正态分布时,则系统的维修度为

$$M(t_1) = \int_0^{t_1} m(t) dt = \int_{-\infty}^{t_1'} m(t') dt' = \int_{-\infty}^{z(t_1')} \phi(z) dz \quad (45.5-20)$$

式中  $t_1' = \ln t_1$ ,

$$z(t_1') = \frac{t_1' - \bar{t}'}{s_t'} \quad (45.5-21)$$

式中的  $\bar{t}'$  和  $s_t'$  见式(45.5-10,11),可以看出,式(45.5-20)中三个等号后的  $m(t)$ 、 $m(t')$  和  $\phi(z)$  分别为对数正态分布、正态分布和标准正态分布。

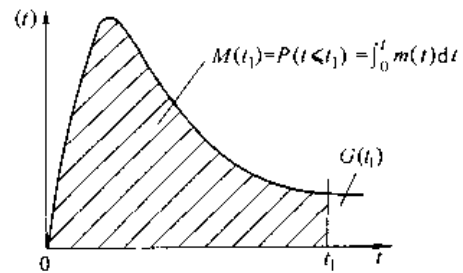


图 45.5-7  $M(t_1)$  与  $m(t)$  的关系

已知  $z(t_1')$ , 由标准正态分布面积表,即可求出维修度  $M(t_1)$ 。

此外,还希望确定所需的修复时间,即在规定的时间内完成维修措施的概率有多大。因此,把式(45.5-20)改写成

$$\begin{aligned} M(t_{1-a}) &= \int_0^{t_{1-a}} m(t) dt = \int_{-\infty}^{t_{1-a}'} m(t') dt' \\ &= \int_{-\infty}^{z(t_{1-a}')} \phi(z) dz \end{aligned} \quad (45.5-22)$$

式中  $z(t_{1-a}') = \frac{t_{1-a}' - \bar{t}'}{s_t'}$  (45.5-23)

常用的维修度值  $M(t_{1-a})$  为 0.80, 0.85, 0.90, 0.95 和 0.99。与此相对应的  $z(t_{1-a}')$  值列于表 45.5-6。

表 45.5-6 与维修度  $M(t_{1-a})$  相应的系数  $z(t_{1-a}')$

维修度 $M(t_{1-a})$	系数 $z(t_{1-a}')$
0.80	0.842
0.85	1.036
0.90	1.282
0.95	1.645
0.99	2.326

由式(45.5-23)

$$t_{1-a}' = \bar{t}' + z(t_{1-a}')s_t'$$

由此可得

$$t_{1-a} = \text{anti ln}(t_{1-a}') = \text{anti ln}[\bar{t}' + z(t_{1-a}')s_t'] \quad (45.5-24)$$

此即当维修度为  $M(t_{1-a})$  时所需的维修时间。

例 已知系统的修复时间数据如表 45.5-3 所示。

①确定维修度函数,并作图表示;②如果维修措施需要在5h内完成,求系统的维修度;③如完成时间为20h,求维修度是多少;④完成维修措施的90%所需的时间;⑤完成维修措施的95%所需的时间。

解 1)当系统中的零件故障时,由式(45.5-20)知,系统的维修度函数为

$$M(t) = \int_{-\infty}^{z(t')} \phi(z) dz$$

由式(45.5-21)

$$z(t') = \frac{t' - \bar{t}'}{s_t}$$

式中  $\bar{t}' = 0.65879, s_t = 1.11435, t' = \ln t$ 。

对于不同的修复时间  $t$ ,可以得出不同的  $M(t)$ ,由此可绘出图 45.5-8。

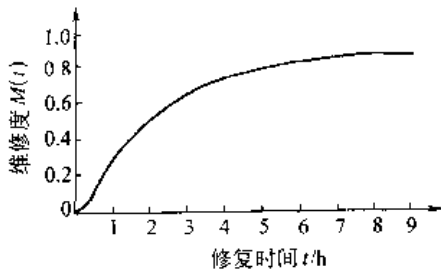


图 45.5-8 维修度函数

2) 修复时间为5h时系统的维修度为

$$M(t = 5h) = \int_{-\infty}^{z(\ln 5)} \phi(z) dz$$

式中  $t' = \ln t = \ln 5 = 1.6094$

$$z(\ln 5) = \frac{t' - \bar{t}'}{s_t} = \frac{1.6094 - 0.65879}{1.11435} = 0.85310$$

因此,由标准正态分布表可得

$$M(t = 5h) = \int_{-\infty}^{0.8531} \phi(z) dz = 0.8032$$

即在5h内完成修复措施的概率是80.32%。

3) 修复时间为20h时系统的维修度为

$$M(t = 20h) = \int_{-\infty}^{z(\ln 20)} \phi(z) dz$$

$$\text{式中 } z(\ln 20) = \frac{2.9957 - 0.65879}{1.11435} = 2.0972$$

因此

$$M(t = 20h) = \int_{-\infty}^{2.0972} \phi(z) dz = 0.9820$$

即在20h内完成维修措施的概率为98.2%。这比在5h内达到的维修度80.32%有很大提高。

4) 由式(45.5-24)和表 45.5-6 可知,完成90%的维修措施的时间为

$$\begin{aligned} t_{0.90} &= \text{antiln}[\bar{t}' + z(t'_{0.90})\sigma_t] \\ &= \text{antiln}(0.65879 + 1.282 \times 1.11435) \\ &= 8.06h \end{aligned}$$

5) 同理,完成95%修复措施的时间,即最大维修时间为

$$\begin{aligned} t_{0.95} &= t_{\text{Max}} \\ &= \text{antiln}(0.65879 + 1.645 \times 1.11435) \\ &= 12.08h \end{aligned}$$

### 2.2.2 已知修复时间为威布尔分布时的维修度函数

当修复时间为威布尔分布时,其概率密度函数为

$$m(t) = \frac{k}{b} \left\{ \frac{t-a}{b} \right\}^{k-1} e^{-\left(\frac{t-a}{b}\right)^k} \quad (45.5-25)$$

式中  $k$ ——形状参数;

$b$ ——尺度参数;

$a$ ——位置参数。

当修复时间为  $t_1$ (h)时的维修度函数为

$$M(t_1) = \int_0^{t_1} m(t) dt \quad (45.5-26)$$

式中  $m(t)$ ——修复时间的概率密度函数。

将式(45.5-25)代入式(45.5-26),可得

$$M(t_1) = 1 - e^{-\left(\frac{t_1-a}{b}\right)^k} \quad (45.5-27)$$

例 已知某设备的修复时间为威布尔分布,其分布参数分别为: $a=0.5h, k=2.5, b=3h$ ,试确定修复时间为5h时的维修度。

解 由式(45.5-27),当  $t_1=5h$  时的维修度为

$$M(t_1 = 5h) = 1 - e^{-\left(\frac{5.0-0.5}{3}\right)^{2.5}} = 0.936$$

即维修措施的93.6%将成功地在5h内完成,如还希望确定所需的修复时间,这时,由式(45.5-27)。可得与已知的维修度  $M(t_1)$  相应的修复时间为

$$t_1 = a + b \{-\ln[1 - M(t_1)]\}^{1/k} \quad (45.5-28)$$

如维修度为  $M(t_1)=0.90$ , 则

$$t_1 = a + b(-\ln 0.1)^{1/k} = a + b(2.302585)^{1/k} \quad (45.5-29)$$

同理,如维修度为  $M(t_1)=0.95$ , 则

$$t_1 = a + b(2.995732)^{1/k} \quad (45.5-30)$$

例 已知某设备的修复时间为威布尔分布,其分布参数分别为: $a=0.5h, k=2.5, b=3h$ ,试确定当维修度  $M(t_1)=0.95$  时所需的修复时间。

解 由式(45.5-30)维修度为  $M(t_1)=0.95$  时的修复时间为

$$t_1 = 0.5 + 3.0(2.995732)^{1/2.5} h = 5.15h$$

即维修措施的95%将在5.15h内完成。

### 2.3 修复率

修复率是修复时间已达到某个时刻但尚未修复的产品,在该时刻后的单位时间内完成修理的概率。表示为



$$\mu(t) = \frac{dM(t)}{dt} \frac{1}{1-M(t)} = \frac{m(t)}{1-M(t)} \quad (15.5-31)$$

式中  $m(t)$ ——修复时间的概率密度函数;  
 $M(t)$ ——维修度。

由式(45.5-18)可知,修复时间的概率密度函数  $m(t)$  是维修度  $M(t)$  的导函数,即

$$m(t) = \frac{dM(t)}{dt} = \frac{d[1-G(t)]}{dt} = -\frac{dG(t)}{dt}$$

式中  $G(t)$ ——不可维修度。

由式(45.5-31),得

$$M(t) = 1 - e^{-\int_0^t m(x)dx} \quad (45.5-32)$$

当为指数分布时,  $\mu(t) = \mu = \text{常数}$ ,故

$$M(t) = 1 - e^{-\mu t} \quad (45.5-33)$$

例 已知系统的修复时间数据如表 45.5-3 所示,

试确定修复率函数,并作图表示之。

解 由式(45.5-12)可得

$$m(t) = \frac{1}{t \times 1.11435 \times \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln t - 0.65829}{1.11435} \right)^2}$$

由式(45.5-20)可得

$$M(t) = \int_0^t m(x)dx = \int_{-\infty}^{\frac{\ln t - 0.65829}{1.11435}} \phi(z)dz$$

当  $m(t)$  及  $M(t)$  已知,代入式(45.5-31)求得  $\mu(t)$  为

$$\mu(t) = \frac{m(t)}{1-M(t)}$$

对于规定的修复时间  $t$ ,可以得出相应的  $m(t)$ 、 $M(t)$  和  $\mu(t)$ ,并将其值列于表 45.5-7,根据表 45.5-7 列出的计算结果,可得修复率函数  $\mu(t)$  的图形如图 45.5-9 所示。

表 45.5-7 维修度  $M(t)$  和修复率  $\mu(t)$  的计算

修复时间 $t$ /h	$t' = \ln t$	$z(t')$	维修度 $M(t)$	修复时间的概率密度函 数 $m(t)$	$1-M(t)$	修复率 $\mu(t)$ 次/h
0.2	-1.6094	-2.0355	0.0210	0.22552	0.9790	0.23035
0.4	-0.9163	-1.4135	0.0788	0.32961	0.9212	0.35780
0.6	-0.5108	-1.0496	0.1469	0.34397	0.8531	0.40320
0.8	-0.2231	-0.7914	0.2144	0.32718	0.7856	0.41646
1.0	0	-0.5912	0.2772	0.30061	0.7228	0.41590
1.2	0.18232	-0.4276	0.3345	0.27227	0.6655	0.40911
1.4	0.33647	-0.2892	0.3862	0.24524	0.6138	0.39954
1.6	0.47000	-0.1694	0.4327	0.22057	0.5673	0.38882
1.8	0.58779	0.0637	0.4746	0.19849	0.5254	0.37780
2.0	0.69315	0.03083	0.5123	0.17892	0.4877	0.36689
2.2	0.78846	0.11636	0.5463	0.16163	0.4537	0.35629
2.4	0.87547	0.19444	0.5771	0.14638	0.4229	0.34611
2.6	0.95551	0.26627	0.6064	0.13290	0.3936	0.33765
3.0	1.09861	0.39469	0.6534	0.11039	0.3466	0.31854
3.6	1.28093	0.55830	0.7117	0.08510	0.2883	0.29517
4.0	1.38629	0.65285	0.7431	0.07232	0.2569	0.28154
4.6	1.52606	0.77827	0.7817	0.05749	0.2182	0.26348
5.0	1.60944	0.85310	0.8032	0.04976	0.1968	0.25281
5.6	1.72277	0.95480	0.8301	0.04053	0.1699	0.23860
6.0	1.79176	1.01671	0.8430	0.03559	0.1570	0.22666
7.0	1.94591	1.15504	0.8760	0.02625	0.1240	0.21161
8.0	2.07944	1.27487	0.8988	0.01986	0.1012	0.19625
9.0	2.19722	1.38057	0.9163	0.01534	0.0837	0.18322
10.0	2.30259	1.47512	0.9299	0.01206	0.0701	0.17205

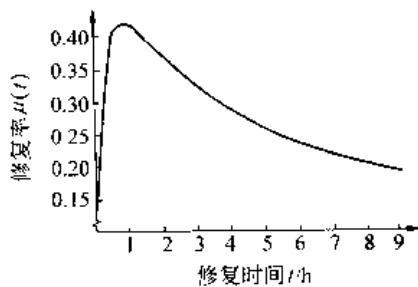


图 45.5-9 修复率函数  $\mu(t)$

## 2.4 平均修复时间

### 2.4.1 主动修复设备的平均时间

在维修性设计中,当仅考虑主动修复时间时,一般是指修复设备的平均时间的稳态值。如设备包括  $N$  个元件,每个元件的 MTBF 分别为  $m_1, m_2, \dots, m_N$ ,当元件 1 失效时,要求平均  $d_1$ h 来修复设备,当元件 2 失效时,要求平均  $d_2$ h 来修复设备,……当元件  $N$  失效时,要求  $d_N$ h。

在运转  $t$  小时后停工的设备,其平均修复时间为

$$MTTR = d = \frac{t}{m_1}d_1 + \frac{t}{m_2}d_2 + \dots + \frac{t}{m_N}d_N$$

$$= t \sum_{i=1}^N \left( \frac{d_i}{m_i} \right) \quad (45.5-34)$$

于是,每一运转小时设备的平均修复时间为

$$d^* = \sum_{i=1}^N \left( \frac{d_i}{m_i} \right) = \frac{d}{t} \quad (45.5-35)$$

若这些元件的失效率为常数,则

$$d = t \sum_{i=1}^N (\lambda_i d_i) \quad (15.5-36)$$

同理

$$d^* = \frac{d}{t} = \sum_{i=1}^N (\lambda_i d_i) \quad (45.5-37)$$

这一时间  $d^*$  应当使之最小,以减少每一运转小时的停工时间;减少事后维修的成本,增加设备的有效度。

### 2.4.2 平均维修工时(MMH)

如元件 1 要求  $n_1$  个维修人员,元件 2 要求  $n_2$  个,……元件  $N$  要求  $n_N$  个,则对于累积运转小时  $t$ ,平均维修工时为

$$MMH = D = \frac{t}{m_1}d_1n_1 + \frac{t}{m_2}d_2n_2 + \dots + \frac{t}{m_N}d_Nn_N$$

$$= t \sum_{i=1}^N \left( \frac{d_i n_i}{m_i} \right) \quad (45.5-38)$$

当  $t=1h$ ,平均维修工时为

$$D^* = \sum_{i=1}^N \left( \frac{d_i n_i}{m_i} \right) \quad (45.5-39)$$

若每个元件的失效率为常数,则

$$D = t \sum_{i=1}^N (\lambda_i d_i n_i) \quad (45.5-40)$$

对于  $t=1h$ ,

$$D^* = \sum_{i=1}^N (\lambda_i d_i n_i) \quad (45.5-41)$$

$D^*$  也叫做每一设备运转小时的平均维修工时。

## 3 按龄期更换的预防维修

按龄期更换的预防维修是维修方法最基本的一种,有下列三种情况,如图 45.5-10 所示。

第一种情况是,只要零件工作满  $T_p h$ ,即使它不失效,也将它更换,如图 45.5-10a 所示。第二种情况是,零件在工作未滿  $T_p h$ , ( $t < T_p$ ) 之前就已失效,并得到事后维修,在事后维修后又工作了  $T_p h$  而未失效,这时,仍将其更换,如图 45.5-10b 所示。第三种情况与第二种情况基本相似,如图 45.5-10c 所示。

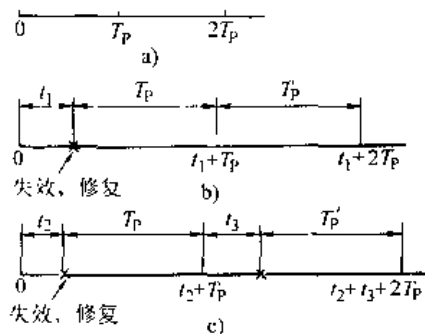


图 45.5-10 按时更换的预防维修

### 3.1 按龄期预防维修设备的可靠度及平均寿命

为了便于分析和预测有预防维修时的可靠度,将时间区间  $[0, t]$  定义如下,并如图 45.5-11 所示。

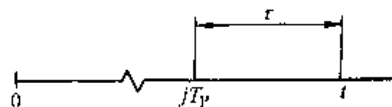


图 45.5-11  $t = jT_p + \tau$

$$t = jT_p + \tau \quad (45.5-42)$$

式中  $j$  — 预防维修的措施次数,  $j=0, 1, 2, \dots$ ;

$\tau$  — 时间变量,  $0 \leq \tau \leq T_p$ ;

$T_p$  — 进行按时更换的预防维修的时间间隔。

如果每  $T_p h$  进行一次预防维修,任务的持续时间为  $t$ ,则产品在  $t$  时刻的可靠度为

$$R_{T_p}(t) = R_{T_p}(jT_p + \tau) = [R(T_p)]^j R(\tau) \quad (45.5-43)$$

式中  $R(T_p)$  — 任务持续时间为  $T_p$  时的可靠度;

$R(\tau)$  — 任务持续时间为  $\tau$  时的可靠度。

由受到这样一系列预防维修的零件所组成的设备或系统的平均寿命为

$$M_{T_p} = \int_0^{\infty} R_{T_p}(t) dt = \sum_{j=0}^{\infty} \int_{jT_p}^{(j+1)T_p} R_{T_p}(t) dt \quad (a)$$

将上式中的积分变量  $t$  改为  $\tau$ , 则有

$$t = jT_p + \tau$$

$$dt = d\tau$$

当  $\tau=0, t=jT_p$ ;

当  $\tau=T_p, t=(j+1)T_p$ 。

于是

$$M_{T_p} = \sum_{j=0}^{\infty} \int_0^{T_p} [R(T_p)]^j R(\tau) d\tau$$

$$= \sum_{j=0}^{\infty} [R(T_p)]^j \int_0^{T_p} R(\tau) d\tau \quad (b)$$

由二项式定理可知,当  $x < 1$ , 有

$$\frac{1}{1-x} = \sum_{j=0}^{\infty} x^j$$

若取  $x=R(T_P)$ , 而  $R(T_P) < 1$ , 则

$$\frac{1}{1-R(T_P)} = \sum_{j=0}^{\infty} [R(T_P)]^j \quad (c)$$

将上式代入式(b), 可得设备或系统的平均寿命为

$$M_{T_P} = \frac{\int_0^{T_P} R(\tau) d\tau}{1-R(T_P)} \quad (45.5-14)$$

只要能够求出  $R(T_P)$  和  $R(\tau)$ , 式(45.5-43)和式(45.5-44)可用于任何设备或系统。

### 3.2 失效时间为威布尔分布时的预防维修

**例** 已知一设备中的轴承的失效时间分布为威布尔分布, 参数如下,  $k=2.0, b=2000h, a=0$ , 试确定下列问题:

- 1) 当没有预防维修时, 在工作小时数分别为 10、100、500 和 1000 时轴承的可靠度;
- 2) 如每隔 10、100 或 500h 对轴承进行一次预防性的更换, 求轴承的可靠度;
- 3) 无预防维修时轴承的平均寿命;
- 4) 每隔 10、100 或 500h 进行一次预防维修时轴承的平均寿命。

**解** 1) 当失效时间为威布尔分布时, 其可靠度函数为

$$R(t) = e^{-\left(\frac{t-a}{b}\right)^k} \quad (45.5-45)$$

将已知参数代入, 可得工作时间  $t$  分别为 10、100、500 和 1000h 时轴承的可靠度为

$$R(t=10h) = e^{-\left(\frac{10}{2000}\right)^2} = 0.999975$$

$$R(t=100h) = e^{-\left(\frac{100}{2000}\right)^2} = 0.999503$$

$$R(t=500h) = e^{-\left(\frac{500}{2000}\right)^2} = 0.939413$$

$$R(t=1000h) = e^{-\left(\frac{1000}{2000}\right)^2} = 0.778801$$

2) 在有按时预防维修时, 由式(45.5-43)和式(45.5-45)可得失效时间为威布尔分布的轴承的可靠度为

$$R_{T_P} = \left[ e^{-\left(\frac{T_P-a}{b}\right)^k} \right]^j e^{-\left(\frac{t-a}{b}\right)^k} \quad (45.5-46)$$

将有关数据代入上式, 得

$$R_{T_P}(t) = e^{-j\left(\frac{T_P}{2000}\right)^2} e^{-\left(\frac{t}{2000}\right)^2}$$

a) 当  $T_P=10h$  时,

$$R_{T_P=10}(t) = e^{-j\left(\frac{10}{2000}\right)^2} e^{-\left(\frac{t}{2000}\right)^2}$$

对于  $0 \leq t \leq 10h, j=0, 1, 2, 3$ , 得到的  $R_{T_P=10}(t)$  值列于表 45.5-8, 并如图 45.5-12 所示。

b) 当  $T_P=100h$  时

$$R_{T_P=100}(t) = e^{-j\left(\frac{100}{2000}\right)^2} e^{-\left(\frac{t}{2000}\right)^2}$$

对于  $0 \leq 100h, j=0, 1, 2, 3$ , 得到的  $R_{T_P=100}(t)$  值列于表 45.5-9, 并如图 45.5-13 所示。

表 45.5-8 当  $T_P=10h$  时的可靠度  $R_{T_P=10}(t)$  的值

$\tau$	$R_{T_P=10}(t)$			
	$j=0$	$j=1$	$j=2$	$j=3$
0	1.0 <sub>8</sub>	0.947500	0.945000	0.942500
1	0.978	0.947498	0.944998	0.942408
3	0.95775	0.947275	0.944775	0.942275
5	0.93375	0.946875	0.944375	0.941875
9	0.947975	0.945475	0.942975	0.940175
10	0.947500	0.945000	0.942500	0.940000

注: 1.0<sub>8</sub> 表示小数点后有 8 个零。

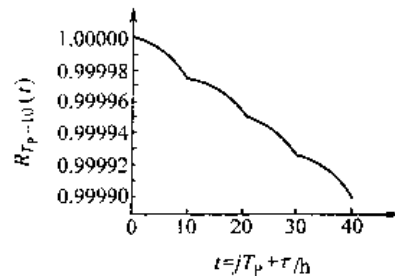


图 45.5-12 每隔 10h 进行一次预防维修得到的可靠度  $R_{T_P=10}(t)$

表 45.5-9 当  $T_P=100h$  时的可靠度  $R_{T_P=100}(t)$  的值

$\tau$	$R_{T_P=100}(t)$			
	$j=0$	$j=1$	$j=2$	$j=3$
0	1.0 <sub>8</sub>	0.99750312	0.99501237	0.99252798
15	0.94378	0.99744692	0.99495653	0.99247225
30	0.9377502	0.99727859	0.99478855	0.99230476
45	0.949388	0.99699814	0.99450880	0.99202571
60	0.9310040	0.99660565	0.99411729	0.99163518
75	0.99859474	0.99610125	0.99361415	0.99113323
90	0.99797705	0.99548510	0.99299954	0.99052021
100	0.99750312	0.99501237	0.99252798	0.99004983

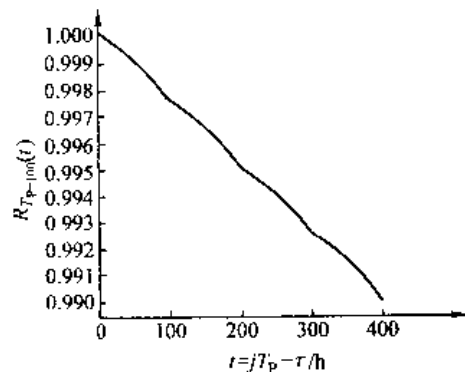


图 45.5-13  $T_P=100h$  时的可靠度  $R_{T_P=100}(t)$

c) 当  $T_p=500h$  时

$$R_{T_p=500}(t) = e^{-j\left(\frac{500}{2000}\right)^2} e^{-\left(\frac{\tau}{2000}\right)^2}$$

对于  $0 \leq \tau \leq 500h, j=0,1,2,3$ , 得到的  $R_{T_p=500}(t)$

值列于表 45.5-10, 并如图 45.5-14 所示。

表 45.5-10 当  $T_p=500h$  时的可靠度  $R_{T_p=500}(t)$  的值

$\tau$	$R_{T_p=500}(t)$			
	$j=0$	$j=1$	$j=2$	$j=3$
0	1.0 <sub>8</sub>	0.93941306	0.88249684	0.82902912
50	0.9337520	0.93882605	0.88193861	0.82851102
100	0.99750312	0.93706740	0.88029252	0.82695902
150	0.99439080	0.93414364	0.87754590	0.82437880
250	0.98449644	0.92484875	0.86881417	0.81617610
300	0.97775124	0.91851222	0.86286156	0.81058413
350	0.96983919	0.91107955	0.85587921	0.80402482
400	0.96078944	0.90257809	0.84789370	0.79652231
450	0.95063509	0.89303896	0.83893252	0.78810466
500	0.93941306	0.88249684	0.82902912	0.77880067

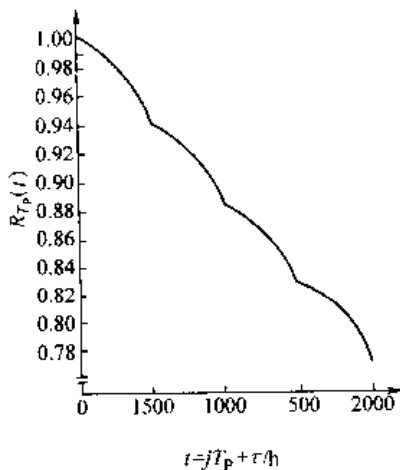


图 45.5-14  $T_p=500h$  时的可靠度

$$R_{T_p=500}(t)$$

由表 45.5-8、表 45.5-9 和表 45.5-10, 可以绘出图 45.5-15。由图可知, 预防维修大大地提高了这些轴承在任务期间内的可靠度。  $T_p$  越小, 则可靠度越大。

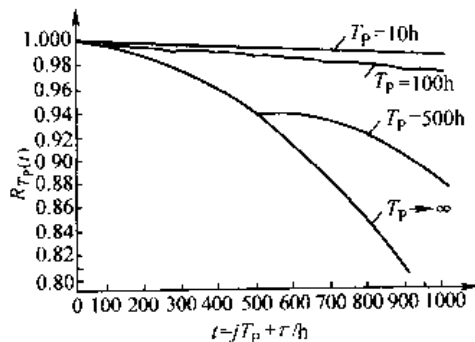


图 45.5-15 预防维修时间与可靠度  $R_{T_p}$  的关系

为了便于比较, 列出了表 45.5-11, 其中第 4 栏为轴承无预防维修时的可靠度, 第 5 栏为每隔  $T_p=100h$  进行预防维修时的可靠度, 第 6 栏为失效时间呈指数分布时的可靠度 (有预防维修时, 失效率  $\lambda_{T_p}=1/M_{T_p}$  为常数)。第 4 栏与第 5 栏的比较结果表明, 有预防维修的轴承其可靠度有了很大改进。第 5 栏与第 6 栏的比较结果表明, 在大约第 2 次更换轴承后, 其可靠度非常接近。

表 45.5-11 无预防维修、每 100h 进行预防更换、以及失效时间为指数分布时的轴承可靠度

任务期 $t/h$	$j$	$\tau$ $/h$	无预防维修时 $R(t) = e^{-\left(\frac{t}{b}\right)^k}$	$T_p=100h$ $R_{T_p}(t) = \zeta R(T_p) \cdot R(\tau)$	$T_p=100h$ $R_{T_p}(t) = e^{-\frac{t}{M_{T_p}}}$
0	0	0	1.0000000	1.0000000	1.0000000
50	0	50	0.999375	0.999375	0.998751
100	1	100	0.997503	0.997503	0.997504
150	1	50	0.994391	0.996880	0.996219
200	1→2	100	0.990050	0.995012	0.995055
250	2	50	0.984496	0.994390	0.993772
300	2→3	100	0.977751	0.992528	0.992531
350	3	50	0.969839	0.991908	0.991292
400	3→4	100	0.960789	0.990050	0.990054
450	4	50	0.950635	0.989431	0.988818
500	4→5	100	0.939413	0.987567	0.987583
700	6→7	100	0.884700	0.982652	0.982659
1000	9→10	100	0.778801	0.975309	0.975320

注: 1.  $b=2000h, k=2.0$ 。

2.  $M_{T_p}$  的求法见式 (45.5-44)。

3) 当失效时间为威布尔分布时, 无预防维修 (即  $T_p=\infty$ ) 时的平均寿命为

$$\bar{i} = a + b\Gamma\left(\frac{1}{k} + 1\right) \quad (45.5-47)$$

式中  $\Gamma$ ——伽玛函数, 由表 45.2.2 查得。

所以, 承的平均寿命为

$$\bar{i} = 2000\Gamma\left(\frac{1}{2} + 1\right) = 1772.46h$$

4) 由式 (45.5-44)。当为威布尔分布时, 每隔  $T_p h$  进行一次预防维修的设备的平均寿命为

$$M_{T_p} = \frac{\int_0^{\infty} R(\tau) d\tau}{1 - R(T_p)} = \frac{\int_0^{\infty} e^{-\left(\frac{\tau}{b}\right)^k} d\tau}{1 - e^{-\left(\frac{T_p}{b}\right)^k}} \quad (45.5-48)$$

a) 当  $T_p=10h$

$$M_{T_p} = \frac{\int_0^{\infty} e^{-\left(\frac{\tau}{2000}\right)^2} d\tau}{1 - e^{-\left(\frac{10}{2000}\right)^2}}$$

分子用 Simpson 法则进行数值积分,得平均寿命

$$M_{T_p} = \frac{9.99994441}{1 - 0.99997500} h = 400003h$$

b) 当  $T_p=100h$

$$M_{T_p=100} = \frac{99.916672914}{1 - 0.99750312} h = 40017h$$

c) 当  $T_p=500h$

$$M_{T_p=500} = \frac{489.775803}{1 - 0.93941305} h = 8084h$$

由图 45.5-15 可以看出,当  $T_p$  减小时,有预防维修的轴承的平均寿命大大增加。因此,对那些失效率随工作时间的增加而增大的零件(这是机械零件的特征,  $k' > 1$ ),通过有规律的、按预先选定的周期进行预防维修,可以得到很大的经济效益。

### 3.3 失效时间为指数分布时的预防维修

(1) 单个元件的预防维修

当无维修时,可靠度为

$$R(t) = e^{-\lambda t} \quad (45.5-49)$$

当每隔  $T_p$  小时进行预防维修时,可靠度为

$$R_{T_p}(t) = R_{T_p}(jT_p + \tau)$$

式中  $t = jT_p + \tau$ , 于是

$$\begin{aligned} R_{T_p}(t) &= [R(T_p)]^j R(\tau) = (e^{-\lambda T_p})^j \times e^{-\lambda \tau} \\ &= e^{-\lambda(jT_p + \tau)} = e^{-\lambda t} = R(t) \end{aligned}$$

这说明,当元件的失效时间为指数分布时,即使在有用寿命期内每  $T_p$  小时更换一次元件,仍不能提高其可靠度。换句话说,在指数分布的有用寿命期内没有必要进行预防维修。

(2) 两并联元件组成的设备的预防维修

当没有维修时,可靠度为

$$R(t) = 2e^{-\lambda t} - e^{-2\lambda t} \quad (45.5-50)$$

当每隔  $T_p$  小时进行预防维修时,可靠度为

$$\begin{aligned} R_{T_p}(t) &= [R(T_p)]^j R(\tau) \\ &= (2e^{-\lambda T_p} - e^{-2\lambda T_p})^j \times (2e^{-\lambda \tau} - e^{-2\lambda \tau}) \end{aligned} \quad (45.5-51)$$

式中  $t = jT_p + \tau$

$$\begin{aligned} j &= 0, 1, 2, 3 \\ 0 &\leq \tau < T_p \end{aligned}$$

例 已知一由两个元件组成的设备的失效率  $\lambda = 0.01$  次/h, 平均寿命 MTBF = 100h, 试确定当  $T_p = 10, 150h$  和  $\infty$  时的可靠度及平均寿命。

解 1) 可靠度计算

a) 当  $T_p = 10h$ , 可靠度为

$$R_{T_p=10}(t) = 2e^{-0.01 \times 10} - e^{-2 \times 0.01 \times 10} \times (2e^{-0.01 \tau} - e^{-0.02 \tau})$$

式中  $0 \leq \tau < 10h$

$$j = 0, 1, 2, \dots, t = j \times 10 + \tau, 0 \leq \tau < \infty$$

b) 当  $T_p = 150h$ , 可靠度为

$$R_{T_p=150}(t) = 2e^{-0.01 \times 150} - e^{-2 \times 0.01 \times 150} \times (2e^{-0.01 \tau} - e^{-0.02 \tau})$$

式中  $0 \leq \tau < 10h$

$$j = 0, 1, 2, \dots, t = j \times 150 + \tau, 0 \leq \tau < \infty$$

c) 当  $T_p \rightarrow \infty$  (无预防维修),  $j = 0$ , 则

$$R_{T_p \rightarrow \infty}(t) = R(t) = 2e^{-0.01 \tau} - e^{-0.02 \tau}$$

式中  $0 \leq \tau < \infty$

对上述可靠度函数进行计算,可以得出与表 45.5-8 同样的表,然后,可据此绘出图 45.5-16。

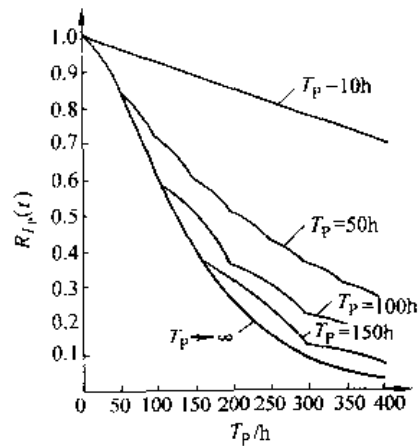


图 45.5-16  $R_{T_p}$  与任务时间关系图

由图 45.5-16 可以看出:①有预防维修的可靠度提高很多。②从一个  $T_p$  到另一个  $T_p$ ,  $R_{T_p}(t)$  是波浪式的,如不采取新的预防维修措施,则  $R_{T_p}(t)$  将下降。③当  $T_p \rightarrow 0$ , 则  $R_{T_p \rightarrow 0}(t) \rightarrow 1$ , 即当设备中的元件不断地更换,则设备将永不失效。④当  $T_p$  较小时 ( $T_p \leq 10h$ ),  $R_{T_p}(t)$  的斜率几乎不变;而当  $T_p$  较大时,则  $R_{T_p}(t)$  的离散程度较大。

2) 平均寿命

当每  $T_p$ h 进行预防维修时,其平均寿命为

$$\begin{aligned} M_{T_p} &= \frac{\int_0^{T_p} R(\tau) d\tau}{1 - R(T_p)} = \frac{\int_0^{T_p} 2(e^{-0.01\tau} - e^{-0.02\tau}) d\tau}{1 - 2e^{-0.01T_p} + e^{-0.02T_p}} \\ &= \frac{150 + 50e^{-0.02T_p} - 200e^{-0.01T_p}}{1 - 2e^{-0.01T_p} + e^{-0.02T_p}} \end{aligned}$$

当  $T_p \rightarrow \infty$ ,  $M_{T_p} = 1.5MTBF = 150h$ ,

$$T_p = 10h, M_{T_p} = 179h,$$

$$T_p = 0h, M_{T_p} \rightarrow \infty$$

将上述的  $T_p$  与  $M_{T_p}$  对应值绘于图 45.5-17 上,可以看出:①当  $T_p$  增大时,  $M_{T_p}$  减小,在开始时  $M_{T_p}$ 。

减小得很快。②当  $T_p=0, M_{T_p} \rightarrow \infty$ 。③当  $T_p \rightarrow \infty, M_{T_p} \rightarrow \infty$  设备无预防维修时的 MTBF。④虽然单个元件的失效率为常数,但两并联元件组成的设备,其失效率不为常数,而是随时间的增加而增加,因此,通过预防维修可以提高其可靠度,即  $R_{T_p}(t) \gg R(t)$ 。

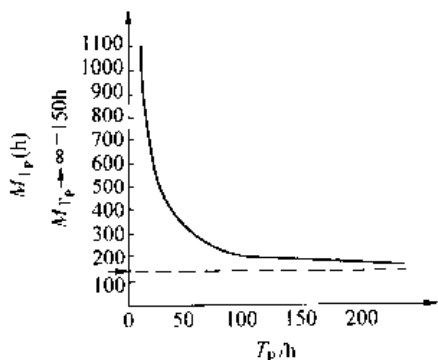


图 45.5-17  $T_p$  与平均寿命  $M_{T_p}$  的关系

#### 4 有贮备的可维修系统的可靠度

有贮备的可维修系统是指一个由  $n$  个单元组成,如其任务期内允许对失效单元进行维修,只要在任务结束前至少还有一个单元在工作的系统。

由  $n$  个单元组成的系统有  $n+1$  个状态:①状态  $n$ —所有的单元都能工作而且在工作;②状态  $n-1$ — $n-1$  个单元能工作而且在工作,1 个单元已经失效并在修理中;③状态  $n-2$ — $n-2$  个单元能工作而且在工作,2 个单元已经失效并在修理中,或者单个修理,或同时修理,……④状态  $1$ —只有一个单元能工作而且在工作,其余的都在修理中,单个修理或同时修理;⑤状态  $0$ —所有单元都已失效。

为了便于分析,作以下假设:①组成系统的各单元的失效时间分布和修复时间分布均为指数分布,即失效率  $\lambda$  和修复率  $\mu$  都为常数;②各单元的失效时间和修复时间是互相独立的;③产生故障的单元经修复后,其失效时间分布与新的单元一样;④在系统开始工作时,各单元都处于正常状态;⑤系统是从状态  $n$  开始工作。

##### 4.1 确定可维修系统可靠度的步骤

1) 弄清楚可维修系统的所有状态,如状态 0、状态 1 等。

2) 写出每一状态在时刻  $t+\Delta t$  时的存在概率  $P_i(t+\Delta t)$ 。只考虑它前一个状态和后一个状态,而不考虑更前面的和更后面的状态,因为其中的阶次较高。

这种方法被称为马尔可夫过程。应注意,所有状态可能存在的概率和必须等于 1,表示为

$$\sum_{i=1}^n P_i(t) = 1 \quad (45.5-52)$$

式中  $P_i(t)$ —状态  $i$  在时刻  $t$  存在的概率;  
 $n$ —状态数。

这些可能存在的状态形成一个完整的集合,它们是互相独立的。

3) 在  $\Delta t$  内失效的概率为  $\lambda\Delta t$ ,因为

$$P(\text{在 } \Delta t \text{ 内失效}) = \frac{N_f \Delta t}{N_0} = \frac{\lambda \Delta t N_0}{N_0} = \lambda \Delta t \quad (45.5-53)$$

式中  $N_f$ —在  $\Delta t$  内的失效数;

$N_0$ — $\Delta t$  开始时的单元数。

因此,在  $\Delta t$  内不失效的概率为  $1 - \lambda\Delta t$ 。同理,在  $\Delta t$  内完成维修和修理措施的概率为  $\mu\Delta t$ ,未完成的概率为  $1 - \mu\Delta t$ 。

4) 将由步骤 3) 推导出的公式列为微分方程。

5) 将由步骤 4) 得出的每一微分方程,用 Laplace 变换求解  $P(t)$ 。

6) 得出  $R_{\lambda,\mu}(t) = 1 - P_0(t)$

式中  $R_{\lambda,\mu}(t)$ —具有失效率为  $\lambda$ 、修复率为  $\mu$  的可维修系统的可靠度。

#### 4.2 两单元并联系统的可靠度

图 45.5-18 所示为一两同样单元组成的并联系统。开始时两个单元都在工作,如果其中一个产生故障,则对其进行修理,一旦修好,立即参加工作。通常,  $\mu > \lambda$ 。

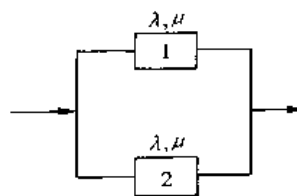


图 45.5-18 两同样单元组成的并联系统

确定这一可维修系统的可靠度,通常有①系统状态分析法;②马尔可夫图解法。

##### 4.2.1 系统的状态分析法

这种方法的步骤为:

1) 弄清楚所有可能存在的状态。在任何时候,系统处于下列三个状态之一,状态 2、状态 1 和状态 0。令  $P_2(t)$ 、 $P_1(t)$  和  $P_0(t)$  分别表示在时刻  $t$  时系统处于状态 2、1 和 0 的概率。由于系统不可能同时存在两种状态,所以这三种状态是互斥事件,因而

$$P_2(t) + P_1(t) + P_0(t) = 1 \quad (45.5-54)$$

2) 写出每一状态在  $t + \Delta t$  时的存在概率  $P_1(t + \Delta t)$ 。令  $P_2(t + \Delta t)$  为在  $t + \Delta t$  时系统处于状态 2 的概率。系统在  $t$  时处于状态 2 的概率为  $P_2(t)$ 。系统的一个单元或另一个单元或两个单元在  $\Delta t$  内失效的概率为  $\lambda\Delta t + \lambda\Delta t - (\lambda\Delta t)^2 - 2\lambda\Delta t - \lambda^2(\Delta t)^2$ 。因  $\Delta t$  很小,  $(\Delta t)^2$  更小, 当  $\Delta t \rightarrow 0$  时, 可忽略不计。所以, 系统在  $t$  时刻处于状态 2 并保持状态 2 而无单元失效的概率为  $P_2(t)(1 - 2\lambda\Delta t)$ 。

系统在  $t$  时刻处于状态 1, 而在  $\Delta t$  内能够修复其中一个失效单元(另一个单元在  $\Delta t$  内未失效)的概率为

$$P_1(t)(\mu\Delta t)(1 - \lambda\Delta t)$$

由于不存在比状态 2 高的状态, 所以, 系统在  $t + \Delta t$  时存在的概率为

$$P_2(t + \Delta t) \approx P_2(t)(1 - 2\lambda\Delta t) + P_1(t)\mu\Delta t - P_1(t)\lambda\mu(\Delta t)^2$$

移项后化简, 可得

$$\frac{P_2(t + \Delta t) - P_2(t)}{\Delta t} = -2\lambda P_2(t) + \mu P_1(t) - \lambda\mu P_1(t)\Delta t$$

当  $\Delta t \rightarrow 0$  时, 对上式取极限, 得

$$P_2'(t) = -2\lambda P_2(t) + \mu P_1(t) \quad (45.5-55)$$

令  $P_1(t + \Delta t)$  为系统在  $(t + \Delta t)$  时处于状态 1 的概率。对于在  $t + \Delta t$  时处于状态 1 的系统, 或者是在  $t$  时处于状态 1、失效的单元未得到修复而工作单元在  $\Delta t$  内未失效, 或是在  $t$  时处于状态 2 而  $\Delta t$  内一个单元失效了。

在  $\Delta t$  内失效单元未在修复的概率为  $1 - \mu\Delta t$ 。在  $\Delta t$  内工作单元未失效的概率为  $1 - \lambda\Delta t$ 。两者同时发生的概率为

$$(1 - \mu\Delta t)(1 - \lambda\Delta t) = 1 - \lambda\Delta t - \mu\Delta t + \lambda\mu(\Delta t)^2$$

由于  $(\Delta t)^2$  的阶次高, 可忽略不计。故上式为

$$(1 - \mu\Delta t)(1 - \lambda\Delta t) = 1 - (\lambda + \mu)\Delta t$$

在  $\Delta t$  内两个工作单元之一失效的概率为

$$\lambda\Delta t + \lambda\Delta t = 2\lambda\Delta t$$

由于系统在  $t$  时处于状态 0 的概率是  $P_0(t)$ , 由状态 0 转移到状态 1 的概率为零(当在状态 0 时不允许修理)。因此, 系统在  $t + \Delta t$  时处于状态 1 的概率为  $P_1(t + \Delta t) = P_1(t)[1 - (\lambda + \mu)\Delta t] + P_2(t)2\lambda\Delta t + P_0(t) \times 0$  移项后化简, 可得

$$\frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} = -(\lambda + \mu)P_1(t) + 2\lambda P_2(t)$$

当  $\Delta t \rightarrow 0$  时, 对上式取极限, 得

$$P_1'(t) = 2\lambda P_2(t) - (\lambda + \mu)P_1(t) \quad (45.5-56)$$

令  $P_0(t + \Delta t)$  为系统在  $t + \Delta t$  时处于状态 0 的概率。系统或在  $t$  时处于状态 0 而且必须处于状态 0, 或

在  $t$  时处于状态 1, 但在  $\Delta t$  内正在工作的单元失效了, 而失效的单元在  $\Delta t$  内未能修复。

系统在时间  $t$  处于状态 0 的概率为  $P_0(t)$ , 而且系统停留在状态 0 的概率为一必然事件, 即为 1。

在  $\Delta t$  内工作单元失效的概率为  $\lambda\Delta t$ , 失效单元未能修复的概率为  $1 - \mu\Delta t$ , 所以

$$P_0(t + \Delta t) = P_0(t) \times 1 + P_1(t)[\lambda\Delta t(1 - \mu\Delta t)]$$

移项后化简, 得

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = \lambda P_1(t) - \lambda\mu\Delta t P_1(t)$$

当  $\Delta t \rightarrow 0$ , 对上式取极限可得

$$P_0'(t) = \lambda P_1(t) \quad (45.5-57)$$

由式(45.5-55~57), 计算系统的可靠度, 得

$$R(t) = 1 - P_c(t) = P_1(t) + P_2(t) \quad (45.5-58)$$

为了求  $R(t)$ , 必须求  $P_0(t)$ 。为此, 作拉氏变换, 在  $t = 0$  时的初始条件为:  $P_2(0) = 1, P_1(0) = 0, P_0(0) = 0$ , 得

$$\begin{aligned} lP_2'(t) = P_2(s) &= \frac{P_2(0)}{s} + \frac{1}{s} \int_0^\infty e^{-st} P_2'(t) dt \\ &= \frac{1}{s} l[P_2'(t)] = \frac{1}{s} P_1'(s) \end{aligned} \quad (45.5-59)$$

式中  $l$ ——拉氏变换算子。

因此

$$l[P_1'(t)] = \int_0^\infty e^{-st} P_1'(t) dt = sP_1(s) \quad (a)$$

同理

$$l[P_0'(t)] = sP_0(s) - 1 \quad (b)$$

$$l[P_0'(t)] = sP_1(s) \quad (c)$$

将式(a)、(b)、(c)代入式(45.5-55)、(45.5-56)和式(45.5-57), 得

$$sP_2(s) - 1 = -2\lambda P_2(s) + \mu P_1(s) \quad (d)$$

$$sP_1(s) = 2\lambda P_2(s) - (\lambda + \mu)P_1(s) \quad (e)$$

$$sP_0(s) = \lambda P_1(s) \quad (f)$$

将式(d)、(e)、(f)重新安排, 得

$$\begin{cases} (s - 2\lambda)P_2(s) - \mu P_1(s) = 1 & (d)' \\ 2\lambda P_2(s) - (\lambda + \mu + s)P_1(s) = 0 & (e)' \\ -\lambda P_1(s) + sP_0(s) = 0 & (f)' \end{cases}$$

解联立方程组, 得

$$P_1(s) = \frac{2\lambda}{s^2 + (3\lambda + \mu)s + 2\lambda^2} \quad (45.5-60)$$

令上式的分母等于 0, 得方程

$$s_1 = \frac{1}{2}[-(3\lambda + \mu) + (\lambda^2 + 6\lambda\mu + \mu^2)^{1/2}] \quad (45.5-61)$$

$$s_2 = \frac{1}{2}[-(3\lambda + \mu) - (\lambda^2 + 6\lambda\mu + \mu^2)^{1/2}] \quad (45.5-62)$$

由式(45.5-60),用部分分数法可得

$$P_1(s) = \frac{2\lambda}{s_1 - s_2} \left( \frac{1}{s - s_1} - \frac{1}{s - s_2} \right) \quad (45.5-63)$$

对  $P_1(s)$  作拉氏逆变换,得

$$P_1(t) = \mathcal{L}^{-1}P_1(s) = \frac{2\lambda}{s_1 - s_2} (e^{s_1 t} - e^{s_2 t}) \quad (45.5-64)$$

由联立方程式(d')、(e')和(f'),可得

$$\begin{aligned} P_2(s) &= \frac{\lambda + \mu + s}{s^2 + (3\lambda + \mu)s + 2\lambda^2} \\ &= \frac{\lambda + \mu + s_1}{(s_1 - s_2)(s - s_1)} - \frac{\lambda + \mu + s_2}{(s_1 - s_2)(s - s_2)} \end{aligned} \quad (45.5-65)$$

对  $P_2(s)$  作拉氏逆变换,得

$$\begin{aligned} P_2(t) &= \mathcal{L}^{-1}P_2(s) \\ &= \frac{(\lambda + \mu + s_1)e^{s_1 t} - (\lambda + \mu + s_2)e^{s_2 t}}{s_1 - s_2} \end{aligned} \quad (45.5-66)$$

同理

$$\begin{aligned} P_0(t) &= \frac{2\lambda^2}{s_1 - s_2} \left( \frac{s_1 - s_2}{s_1 s_2} + \frac{s_2 e^{s_1 t} - s_1 e^{s_2 t}}{s_1 s_2} \right) \\ &= 1 + \frac{s_2 e^{s_1 t} - s_1 e^{s_2 t}}{s_1 - s_2} \end{aligned} \quad (45.5-67)$$

由式(45.5-58),得可维修系统的可靠度为

$$R(t) = 1 - P_0(t) = \frac{s_1 e^{s_2 t} - s_2 e^{s_1 t}}{s_1 - s_2} \quad (45.5-68)$$

#### 4.2.2 马尔可夫图解法

马尔可夫图解法如图 45.5-19 所示。图中小的圆圈称为节点,它代表一种系统状态。每个节点上的回路称为自我回路,上面表示节点在  $\Delta t$  内保持这一状态的概率。每个节点表示一种状态,例如  $x_1, \bar{x}_2$  表示单元 1 成功和单元 2 失效。节点之间由直线和曲线联接起来。直线表示由于失效而状态转移,曲线表示由于修复而状态转移。在每一线路上写明了转移概率。每一转移概率的系数称为转移系数。在绘制好马尔可夫图之后,为了求系统可靠度,必须注意:

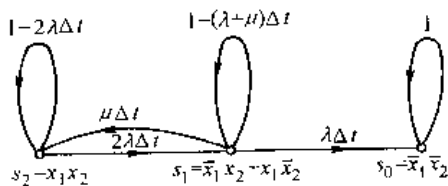


图 45.5-19 有维修的两相同单元

并联系统的马尔可夫图

1) 离开每一节点的线路上的转移概率之和等

于 1。

2) 在  $t + \Delta t$  时任一节点(状态)存在的概率等于所有到达该节点的信号的和。所有其他节点被认为是时间  $t$  时的概率源。

3) 每一系统状态的微分方程的算法如下:

① 使任一节点上概率的导数等于进入节点的转移概率的和。② 使自我回路上任一整数增益系数为零。③ 不考虑  $\Delta t$  系数。

按照上述算法,由图 45.5-19 可得出下列各式

$$\begin{cases} P_2'(t) = -2\lambda P_2(t) + \mu P_1(t) \\ P_1'(t) = 2\lambda P_2(t) - (\lambda + \mu) P_1(t) \\ P_0'(t) = \lambda P_1(t) \end{cases}$$

很明显,此即式(45.5-55~57)。因此,可按前述的方法求出系统的可靠度。

例 由两相同单元组成的可维修并联系统,已知其  $\lambda = 0.01$  次/h,  $\mu = 1.0$  次/h 试确定:

- 1) 系统的可靠度是多少?
  - 2) 如果没有维修,可靠度是多少?
  - 3) 如果只是—一个单元而无维修,可靠度是多少?
- 用图、表来比较三种结果。

解 计算结果见表 45.5-12。

表 45.5-12 计算结果

任务期 $t/h$	(a) $R(t) = \frac{s_1 e^{s_2 t} - s_2 e^{s_1 t}}{s_1 - s_2}$	(b) $2e^{-\lambda t} - e^{-2\lambda t}$	(c) $e^{-\lambda t}$
50	0.990524	0.845182	0.606531
100	0.980952	0.600423	0.367879
150	0.971473	0.396473	0.223130
200	0.962086	0.252355	0.135335
250	0.952789	0.157432	0.082084
300	0.943583	0.097095	0.049787
400	0.925435	0.036296	0.018315
500	0.907636	0.013430	0.006737
600	0.890180	0.004951	0.002479
700	0.873060	0.001823	0.000919
800	0.856260	0.000670	0.000335
900	0.839800	0.000246	0.000123
1000	0.823649	0.000090	0.000045

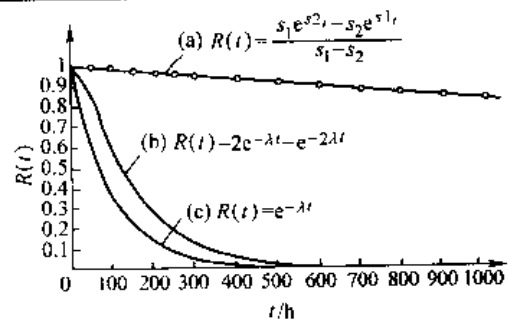


图 45.5-20 系统的可靠性曲线

(a) 两相同单元有维修的并联系统 (b) 两相同单元无维修的并联系统 (c) 一单元无维修系统



上述三种情况下的系统可靠度,如图 45.5-20 所示,从中可以看出,有维修与无维修,有贮备与无贮备之间的明显差别。

### 5 可维修系统的有效度

#### 5.1 有效度的定义

系统的有效度可以用下列三种方式来定义。

1) 瞬时有效度  $A(t)$  产品在某时刻具有或维持其规定功能的概率。换句话说, $A(t)$ 是系统在运行后的任何一个随机时间  $t$  时可以使用的概率。

2) 平均有效度  $A_m(t_1, t_2)$ ——在某个规定时间区间  $[t_1, t_2]$  内有效度的平均值,也叫任务有效度。

3) 稳态有效度  $A_s$  当时间趋于无限时,瞬时有有效度的极限值,也叫极限有效度,表示为

$$A_s = \lim_{t \rightarrow \infty} A(t)$$

#### 5.2 可维修的单个部件的有效度

图 45.5-21 所示为一可维修的单个部件。它有两种可能的状态:

状态 1——部件在工作,因而可以使用。

状态 0——部件已经失效,并正在修理。

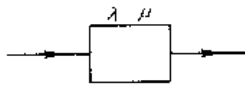


图 45.5-21 可维修的单个部件

该部件的马尔可夫图如图 45.5-22 所示。

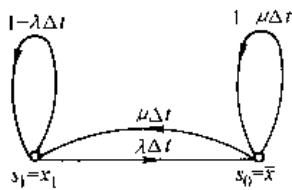


图 45.5-22 可维修的单个部件的马尔可夫图

单个部件的有效度为

$$A(t) = P_1(t) = 1 - P_0(t) \quad (45.5-69)$$

式中  $P_0(t)$  ——部件的不有效度。

由马尔可夫图可得下列微分方程

$$\begin{cases} P_1'(t) = -\lambda P_1(t) + \mu P_0(t) \\ P_0'(t) = \lambda P_1(t) - \mu P_0(t) \end{cases}$$

初始条件为:  $P_1(0) = 1, P_0(0) = 0$ 。作拉氏变换,

得

$$\begin{cases} (s + \lambda)P_1(s) - \mu P_0(s) = 1 \\ -\lambda P_1(s) + (s + \mu)P_0(s) = 0 \end{cases}$$

联立求解,得

$$P_1(s) = \frac{\mu}{s(s + \lambda + \mu)} + \frac{\mu}{s(s + \lambda + \mu)} - \frac{1}{s + \lambda} \frac{1}{\mu} + \frac{\mu}{s_1 - s_2} \left( \frac{1}{s - s_1} - \frac{1}{s - s_2} \right) \quad (45.5-70)$$

式中  $s_1 = 0, s_2 = -(\lambda + \mu)$ 。

式(45.5-70)可写为

作拉氏逆变换,得

$$P_1(t) = 1 - P_1(s) = e^{-(\lambda + \mu)t} + \frac{\mu}{\lambda + \mu} [1 - e^{-(\lambda + \mu)t}] \quad (45.5-71)$$

所以部件的瞬时有有效度为

$$A(t) = P_1(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \quad (45.5-72)$$

式中第一项为稳态元件,第二项为转移元件。

该部件的任务有效度为

$$A_m(t_2 - t_1) = \frac{t}{t_2 - t_1} \int_{t_1}^{t_2} A(t) dt \quad (45.5-73)$$

稳态有效度为

$$A_s = \lim_{t \rightarrow \infty} A(t) = A(\infty) = \frac{\mu}{\lambda + \mu} + \frac{1}{1 + \frac{\lambda}{\mu}} \quad (45.5-74)$$

当为指数分布时,  $\lambda = \frac{1}{MTBF}, \mu = \frac{1}{MTBR}$ , 则稳态

有效度为

$$A_s = \frac{MTBF}{MTBF + MTTR} \quad (45.5-75)$$

通常,因为修复率  $\mu$  比失效率  $\lambda$  大,  $\lambda/\mu$  为一分数,所以由式(45.5-74),稳态有效度可以表示为

$$A_s = \frac{1}{1 + \frac{\lambda}{\mu}} = 1 - \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2} - \frac{\lambda^3}{\mu^3} + \dots \approx 1 - \frac{\lambda}{\mu} \quad (45.5-76)$$

图 45.5-23 为可维修的单个部件的有效度。图中表现了瞬时有有效度  $A(t)$  怎样作为  $\frac{t}{\lambda + \mu}$  的函数而趋近于稳态有效度。

由图还可看出,  $A(t)$  分为过渡(瞬态)部分和稳定部分,在  $t \approx \frac{4}{\mu + \lambda}$  之前,过渡(瞬态)部分下降得比较快而可以忽略不计。如果  $\mu \gg \lambda$ , 则瞬态部分在  $t \approx \frac{4}{\mu}$  之前就可以忽略不计。

例 已知可维修的单一部件的失效率  $\lambda = 0.01$  次/h, 修复率  $\mu = 1.0$  次/h, 求:

1) 工作时间为 0.25、0.5、0.75、1、1.5、2、2.5、3、3.5、4、5、6、7、8、9、10h 时的瞬时有有效度;

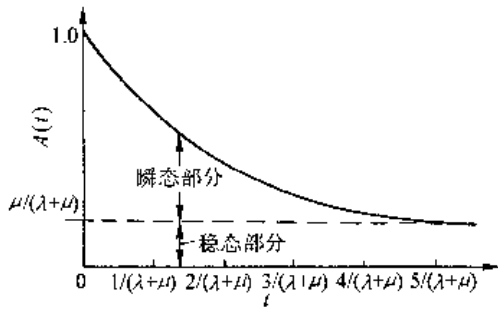


图 45.5-23 可维修的单个部件的有效度  
2) 稳态有效度。并将结果列成表和绘成图。

解 1) 由式(45.5-72),瞬态有效度为

$$A(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}$$

所以

$$\begin{aligned} A(t) &= 0.25h \\ &= \frac{1}{0.01 + 1} + \frac{0.01}{0.01 + 1} e^{-(0.01 + 1) \times 0.25} \\ &= 0.997791 \end{aligned}$$

同理,可得计算结果如表 45.5-13。

表 45.5-13 计算结果

工作时间 t/h	瞬时有效度 A(t)	工作时间 t/h	瞬时有效度 A(t)
0	1.000000	3.50	0.990388
0.25	0.997791	4.00	0.990273
0.50	0.996074	5.00	0.990162
0.75	0.994741	6.00	0.990122
1.00	0.993705	7.00	0.990107
1.50	0.992275	8.00	0.990102
2.00	0.991412	9.00	0.990100
2.50	0.990892	10.00	0.990099
3.00	0.990577	∞	0.990099

2) 由式(45.5-74),稳态有效度为

$$A_s = A(\infty) = \frac{\mu}{\lambda + \mu} = \frac{1}{0.01 + 1} = 0.99099$$

上述计算结果绘于图 45.5-24。

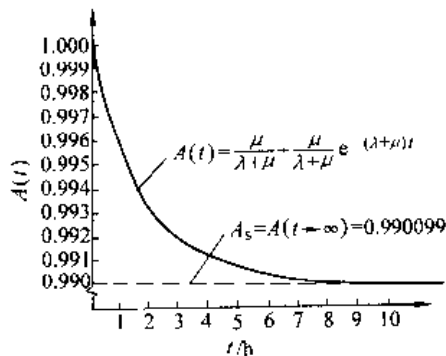


图 45.5-24 可维修的单个部件  
的有效度(λ=0.01, μ=1)

### 5.3 可维修的两单元并联系统的有效度

图 45.5-25 为一可维修的由两相同单元组成的并联系统。它所处的状态可能有以下三种:

状态 2——两个单元都在工作;

状态 1——一个单元在工作,另一个已发生故障并正在修理;

状态 0——两个单元都已发生故障,并同时正在修理。

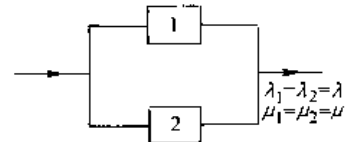


图 45.5-25 可维修的由两  
相同单元组成的并联系统

对于两个故障单元进行同时修理的情况,其马尔可夫图如图 45.5-26 所示。

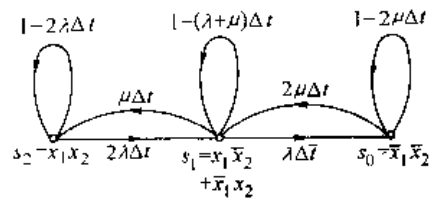


图 45.5-26 两相同单元并联系  
统同时修理的马尔可夫图

由系统所处的状态,可写出下列的微分方程

$$\begin{cases} P_2'(t) = -2\lambda P_2(t) + \mu P_1(t) \\ P_1'(t) = 2\lambda P_2(t) - (\lambda + \mu)P_1(t) + 2\mu P_0(t) \\ P_0'(t) = \lambda P_1(t) - 2\mu P_0(t) \end{cases}$$

初始条件为:  $P_2(0) = 1, P_1(0) = P_0(0) = 0$ 。作拉氏变换得

$$\begin{cases} (s + 2\lambda)P_2(s) - \mu P_1(s) = 1 \\ -2\lambda P_2(s) + (s + \lambda + \mu)P_1(s) - 2\mu P_0(s) = 0 \\ -\lambda P_1(s) + (s + 2\mu)P_0(s) = 0 \end{cases}$$

联立求解,得

$$P_0(s) = \frac{2\lambda^2}{s[s^2 + (3\lambda + 3\mu)s + (2\lambda^2 + 2\mu^2 + 4\lambda\mu)]}$$

由上式中的分母,可得

$$s_1 = -2(\lambda + \mu), s_2 = -(\lambda + \mu), s_3 = 0$$

由部分系数法,可得

$$P_0(s) = \frac{2\lambda^2}{s_1 s_2^2} + \frac{2\lambda^2}{s_1(s_1 - s_2)(s - s_1)} - \frac{2\lambda^2}{s_2(s_1 - s_2)(s - s_2)}$$

(45.5-77)

作拉氏逆变换,得

$$P_0(t) = \frac{2\lambda^2}{s_1 s_2^2} + \frac{2\lambda^2}{s_1(s_1 - s_2)} e^{s_1 t} - \frac{2\lambda^2}{s_2(s_1 - s_2)} e^{s_2 t} \quad (45.5-78)$$

系统的瞬态有效度为

$$A(t) = 1 - P_0(t) = \frac{\mu^2 + 2\lambda\mu}{\lambda^2 + 2\lambda\mu + \mu^2} + \frac{2\lambda^2}{s_1 s_2 (s_1 - s_2)} (s_1 e^{s_1 t} - s_2 e^{s_2 t}) \quad (45.5-79)$$

式中第 1 项为稳态部分,第 2 项为瞬态部分。系统的稳态有效度为

$$A_\infty = A(\infty) = \frac{\mu^2 + 2\lambda\mu}{\lambda^2 + 2\lambda\mu + \mu^2} = 1 - \left( \frac{\lambda}{\lambda + \mu} \right)^2 \quad (45.5-80)$$

图 45.5-27 描绘了有效度、修复率、失效率之间的关系,并说明了在能够多路维修的情况下,系统的有效度得以提高。

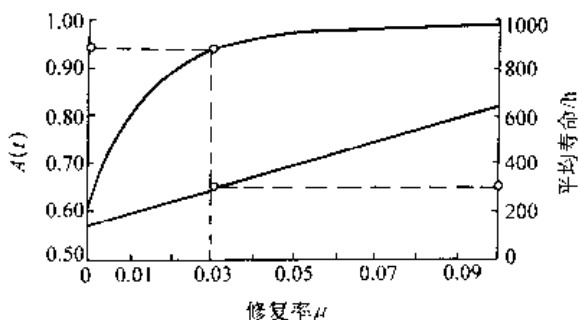


图 45.5-27 可维修的情况下系统有效度的提高

### 6 维修性设计例子

某厂生产的离心式水泵采用按固定时间间隔的预防维修,已知预防维修和事后维修的工作费用为 60 元/h,停工损失为 100 元/h,子样容量  $n=9$ ,如果目标可靠度为 0.99,置信度为 0.90,试求

- 1) 确定零件的维修分组;
- 2) 确定与一定的预防维修时间  $T_P$ ,相应的零件可靠度;
- 3) 确定每一零件的维修费用公式;
- 4) 对每一分组,优化其预防维修时间;
- 5) 优化整个系统的预防维修时间。

解 1) 为了确定零件在维修时的分组,需先确定其更换时间  $t_0$ ,为防止零件进入其寿命期的耗损阶段。 $t_0$  是平均耗损时间  $\bar{t}_w$  和标准差  $\sigma_1$  的函数,表示为

$$t_0 = \bar{t}_w - k\sigma_1 \quad (45.5-81)$$

- 式中  $\bar{t}_w$  ——平均耗损寿命(h);  
 $\sigma_1$  ——失效时间的标准差(h);  
 $k$  ——标准变量。

由于计算是根据有限个子样进行的,因此要利用估计值  $\bar{t}_w$  和子样的标准差  $s$ ,如表 45.5-14 所示。

表 45.5-14 某型号水泵主要零件的  $\bar{t}_w$  和  $s(n=9)$

主要零件	平均耗损时间 $\bar{t}_w$ /h	失效时间的标准差 $s$ /h
壳体	60000	6000
转子	20000	2500
轴	12000	1500
轴套	8000	1000
轴承	12000	1833

为了偏于安全,使用  $t_{WL2}$  和  $s_{U1}$  来估计水泵零件的更换时间,所以

$$t_0 = t_{WL2} - K s_{U1} \quad (45.5-82)$$

式中  $t_{WL2}$  ——平均耗损时间的双侧置信区间下限;  
 $s_{U1}$  ——平均失效时间的标准差的单侧置信区间上限;

$K$  ——标准变量。

平均耗损时间的双侧置信区间下限为

$$t_{WL2} = \hat{t}_w - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \quad (45.5-83)$$

式中  $t_{\frac{\alpha}{2}, n-1}$  —— $t$  分布的分位点,因为子样容量较小,所以不用正态分布。

置信度  $C=1-\gamma=0.90$ 。

将有关数据代入上式,得壳体的

$$t_{WL2} = 60000 - t_{0.05, 8} \frac{6000}{\sqrt{9}} = 56280\text{h}$$

因可靠度的目标值  $R_0=0.99, K=2.33$ 。

$$s_{U1} = s \left( \frac{n-1}{s_1^2 - s_{1, n-1}^2} \right) = 6000 \left( \frac{9-1}{3.49} \right) \text{h} = 9078\text{h}$$

式中  $s_{0.99, 8}^2 = 3.49$ 。

因此,水泵壳体的更换时间为

$$t_0 = (56280 - 2.33 \times 9078)\text{h} = 35128\text{h}$$

同理,水泵转子的更换时间为

$$t_0 = 9638\text{h}$$

轴的更换时间为

$$t_0 = 5783\text{h}$$

轴套的更换时间为

$$t_0 = 3855\text{h}$$

轴承的更换时间为

$$t_0 = 4403\text{h}$$

在得出这 5 个主要零件的更换时间之后,对零件进行分组。第 1 组是预防维修时间低于 6000h 的零件,第 2 组的预防维修时间为 6000~10000h,第 3 组的预防维修时间大于 10000h。

表 45.5-15 列出了三个分组的所有零件表,其中还包含了每个零件的近似价格以及平均普通失效率  $\lambda_{\text{普通}}$ 。这些数据可以用于维修费用的优化。

2) 与一定的预防维修时间  $T_P$  相应的零件可靠度如表 45.5-16 所示。

表 45.5-15 某型号离心式水泵所有零件的价格及失效率  $\lambda_{gen}$

零件序号	第 1 组		价格/零件 (元)	平均普通失效率 $\lambda_{gen} 10^{-5}$ 次/h
	名称	数目要求		
1	轴	1	50	0.35
2	轴套	2	25	0.30
10	1号球轴承	1	50	1.57
17	2号球轴承	1	50	1.34
		共计 5		
第 2 组				
24	转子	1	100	0.15
第 3 组				
3	轴螺母(右旋)	1	1	0.0042
4	轴螺母(左旋)	1	1	0.0042
5	壳体环	2	2.50	0.10
6	壳体衬套	2	5	0.1084
7	轴承衬套	1	5	0.137
8	衬套盖	1	2	0.012
9	隔套	1	5	0.05
11	轴承压板	1	5	0.088
12	轴承锁紧垫圈	1	0.5	0.002
13	轴承锁紧螺母	1	1	0.002
14	轴承衬套	1	5	0.137
15	衬套盖	1	2	0.012
16	隔套	1	5	0.05
18	轴承锁紧垫圈	1	0.5	0.002
19	轴承锁紧螺母	1	1	0.002
20	注油嘴	2	0.5	0.055
21	平键	1	5	0.14
22	键	1	1	0.09
23	O型环	1	1	0.04
25	转子环	1	5	0.10
26	螺钉	12	0.42	0.025
27	锁紧垫圈	12	0.42	0.025
28	见注			
29	壳体(上半部)	1	300	--
30	壳体(下半部)	1	425	--
31	轴承端部平板	1	5	0.088
32	压盖(剖分式)	4	5	0.125
33	压盖螺栓	4	0.50	0.034
34	密封罩(剖分式)	4	3.15	0.05
35	油孔盖	2	0.50	0
36	轴环	2	1.00	0
37	阀杆	2	2.50	0.336
38	衬垫	1	2	0.700
39	真空衬垫	1	5	0.138
40	放气衬垫	1	5	0.138
41	直榫	2	5	0.008
42	管塞	1	2	0
43	排气旋塞手柄	1	2	0.084
44	六角螺母	4	0.50	0.016
45	螺钉	20	0.80	0.04
46	壳体衬垫	1	5	0.138
47	六角定位螺钉	2	0.05	0.0084
48	曲轴箱密封	1	5	0.01
49	轴承盖	1	2	0.2666
50	0号斜棒	1	3	0.016
51	1号斜棒	1	3	0.008
52	轴承盖螺钉	1	1	0.044
53	联轴器锁紧螺母	1	1	0.04
		共计		

注:回转部件——主要包括下列零件:1,2,3,4,8,9,10,11,12,13,15,17,18,19,20,21和24。

表 45.5-16 对于选定的预防维修时间  $T_p$ , 轴的可靠度

预防维修时间 $T_p$ /h	偶然失效期内的可靠度 $R_c(t) = e^{-\lambda t}$	耗损失效期内的可靠度 $R_w(t) = \int_0^t \frac{12000}{.005} \phi(x) dx$	轴的可靠度 $R(t) = R_c(t)R_w(t)$
0	1	1	1
1000	0.9965	1	0.9965
2000	0.9930	1	0.9930
3000	0.9896	1	0.9896
4000	0.9861	1	0.9861
5000	0.9827	1	0.9827
6000	0.9792	1	0.9792
7000	0.9758	0.9996	0.9754
8000	0.9725	0.9962	0.9688
9000	0.9690	0.9773	0.9470
10000	0.9656	0.9083	0.8770
11000	0.9623	0.7486	0.7204
12000	0.9589	0.5000	0.4795
13000	0.9555	0.2514	0.2402
14000	0.9522	0.0917	0.0873
15000	0.9489	0.0227	0.0215
16000	0.9455	0.0038	0.0036
17000	0.9423	0.0004	0.0004
18000	0.9390	0.0000	0.0000

注:由表 45.5-14 轴的平均耗损时间为 12000h, 标准差为 1500h。

同理,可以计算出第 1 组内其他零件(轴套、轴承)的可靠度  $R(t)$ 。这些零件的可靠度与预防维修时间  $T_p$  之间的关系,如图 45.5-28 所示。由于转子是第 2 组内唯一的零件,其可靠度  $R(t)$  与  $T_p$  的关系也绘于同一图中。

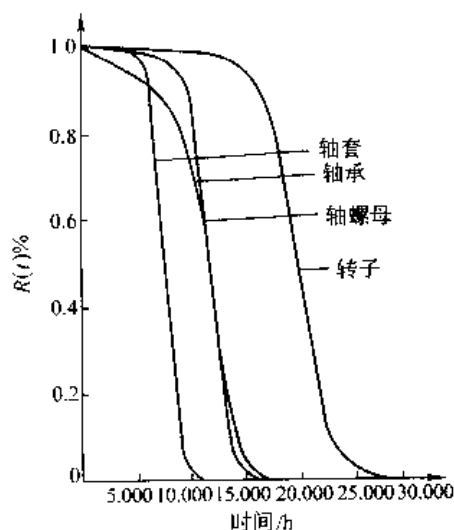


图 45.5-28 第 1 组、第 2 组零件  $R(t)$  与  $T_p$  的关系

第 3 组零件只需进行事后维修。

$$t = (0.2 + 0.6 + 0.2 + 0.2 + 0.2 + 0.1)h$$

3) 拆卸水泵所需要的工时如表 45.5-17 所示,由

$$+ (0.2 + 0.2 + 0.8 + 0.8 + 0.6 + 0.05 + 0.05)h$$

表 45.5-17 可知,拆卸轴所需的时间为

$$= 4.2h$$

表 45.5-17 拆卸水泵所需要的工作

拆卸零件	电机	上盖	环	轴承盖	回转部件	键	
所需工时/h	0.2	0.6	0.2	0.2	0.2	0.1	
A. 拆卸一个轴承							
拆卸零件	轴承盖螺钉	轴承衬套盖	轴承衬套	轴承锁紧螺母	锁紧垫圈	球轴承	
所需工时/h		0.1			0.1		
B. 拆卸轴套							
拆卸零件	轴承压板	隔套	轴螺母	壳体衬套	轴套		
所需工时/h	0.1	0.2	0.2	0.2	0.1		
C. 拆卸转子							
拆卸零件	A	B	转子				
所需工时/h	0.2	0.8	0.6				
D. 拆卸轴							
拆卸零件	拆卸一个轴承 (近离合器一端)	拆卸一个轴承 (另一端)	拆卸轴套(近 离合器一端)	拆卸轴套 (另一端)	转子	离合器	轴
所需工时/h	0.2	0.2	0.8	0.8	0.6	0.05	0.05

因此,轴的更换需要 8.4h。

考虑到在预防维修时间间隔内可能会有事后维修,设备每小时维修费用为

$$C_T = C_{PM} + C_{CM} = \frac{C_{Pj}}{T_P} + \frac{C_{Cj}}{m_j(T_P)} \quad (45.5-84)$$

式中  $C_{PM}$ ——预防维修费用;

$C_{CM}$ ——事后维修费用;

$C_{Pj}$ ——第  $j$  个零件的预防维修费用;

$C_{Cj}$ ——第  $j$  个零件的事后维修费用;

$T_P$ ——预防维修时间间隔;

$M_j(T_P)$ ——有预防维修的第  $j$  个零件的 MTBF。

$$\frac{1}{m_j(T_P)} = \lambda_c - \frac{T_P \int_0^{T_P} R(t) dt}{T_P \int_0^{T_P} R(t) dt} \quad (45.5-85)$$

已知轴的成本为 50 元(表 45.5-15),维修工人的工资为 6 元/h,停工损失为 100 元/h,由式(45.5-84)轴的维修费用  $C_{sh}$  为

$$C_{sh} = \frac{8.4 \times 6 + 50}{T_P} + \frac{8.4(6 + 100)50}{m_j(T_P)} = \frac{100.4}{T_P} + \frac{940.4}{m_j(T_P)} \quad (45.5-86)$$

由上式,可以求出不同预防维修时间  $T_P$  时轴的维修费用

例如,当  $T_P = 10000h$ , 则  $C_{PM} = \frac{100.4}{10000} h =$

0.01004h 时,由式(45.5-85),轴的失效率为

$$\lambda = \frac{1}{m_j(10000)} = \frac{10000 - \int_0^{10000} R(t) dt}{10000 \times \int_0^{10000} R(t) dt}$$

由表 45.5-16 知,  $R(t) = 0.8770$ , 所以

$$\lambda_c = \frac{1}{m_j(10000)} = \frac{10000 - 8770}{10000 \times 8770} \text{次/h} = 14.03 \times 10^{-6} \text{次/h}$$

由式(45.5-86)

$$C_{CM} = 940.4 \times 14.03 \times 10^{-6} \text{元/h} = 0.01319 \text{元/h}$$

因此,如果水泵每工作 10000h 就更换一次轴,则预防维修加事后维修的总费用为

$$C_{sh} = 0.01004 + 0.01319 \text{元/h} = 0.02323 \text{元/h}$$

对于不同的预防维修时间  $T_P$ , 轴的维修费用列于表 45.5-18。

表 45.5-18 对于选定的预防维修时间  $T_P$ ,

第 1 组零件的维修费用

预防维修时间 $T_P$ /h	轴的维修费 元/h	轴套的维修费 元/h	1号轴承的维修费 元/h	2号轴承的维修费 元/h
5000	0.02337	0.01221	0.01977	0.02079
6000	0.02007	0.01235	0.01748	0.01852
7000	0.01773	0.02314	0.01602	0.01707
8000	0.01634	0.07373	0.01552	0.01652
9000	0.01699	0.31801	0.01641	0.01754
10000	0.02323	2.28377	0.02041	0.02169

同理,对于每一组的其他零件,如轴套和轴承的维修费用,分别用  $C_{sj}$  和  $C_{bk}$  表示,用公式表示如下

$$C_{sj} = \frac{52.6}{T_P} + \frac{512.6}{m_{sj}(T_P)} \quad (45.5-87)$$

$$C_{be} = \frac{70.4}{T_P} + \frac{410.4}{m_{be}(T_P)} \quad (45.5-88)$$

对于选定的  $T_P$ , 这些零件的维修费用也可列于表 45.5-18 中。

第 2 组内唯一的零件转子的维修费用公式为

$$C_{r0} = \frac{137.2}{T_P} - \frac{757.2}{m_{r0}(T_P)} \quad (45.5-89)$$

4) 为了优化第 1 组的预防维修时间  $T_P$ , 将第 1 组的 5 个零件(其中有 2 个轴套)的维修费用加在一起, 可得不同预防维修时间  $T_P$  时的第 1 组维修总费用如表 45.5-19 所示。

表 45.5-19 不同  $T_P$  时, 第 1 组维修总费用

$T_P/h$	第 1 组零件维修总费用(元/h)
5000	0.08836
6000	0.08077
7000	0.09710
8000	0.14584
9000	0.68696
10000	4.63287

根据上列数据, 可得第 1 组零件的维修总费用与预防维修时间  $T_P$  的关系曲线, 如图 45.5-29 所示。

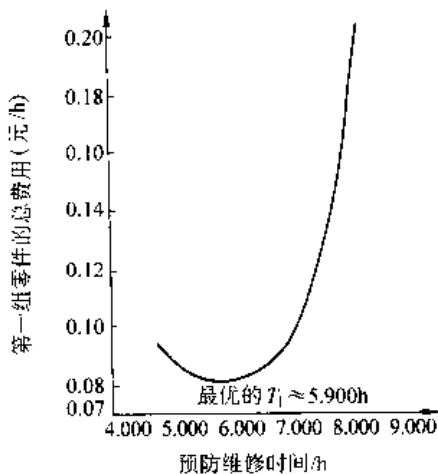


图 45.5-29 第 1 组零件的维修总费用与预防维修时间的关系

由图 45.5-29 可知, 最优的  $T_{P1} = 5900h$ , 总费用为  $C_T = 0.08050$  元/h。同理, 第 2 组零件的维修总费用与预防维修时间  $T_P$  的数据见表 45.5-20。

表 45.5-20 第 2 组维修总费用

$T_P/h$	第 2 组零件维修总费用(元/h)
10000	0.01486
11000	0.01362
12000	0.01262
13000	0.01185
14000	0.01140
15000	0.01150
16000	0.01254
17000	0.01516

根据上列数据, 可得第 2 组零件的维修总费用与预防维修时间  $T_P$  的关系曲线, 如图 45.5-30 所示。

由图 45.5-30 可知, 最优的  $T_{P2} \approx 14350h$ , 总费用为  $C_T = 0.01138$  元/h。

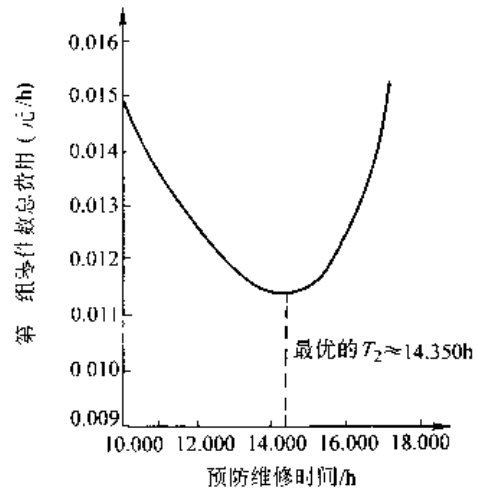


图 45.5-30 第 2 组零件的维修总费用与预防维修时间的关系

5) 不难看出, 由图 45.5-29 和图 45.5-30 得出的最佳预防维修时间  $T_{P1}$  和  $T_{P2}$  分别大于按照公式 (45.5-82) 所得的更换时间  $t_0$  (轴与转子的  $t_0$  分别为 5783h 与 9638h)。但上述的  $t_0$  值是按置信度  $= 0.90, R(t) = 0.99$  确定的, 在求得  $T_{P1} = 5900h$  和  $T_{P2} = 14350h$  之后, 意味着同意接受的可靠度见表 45.5-21。

表 45.5-21 同意接受的可靠度

零件	可靠度
轴	0.9795
轴套	0.9649
1号轴承	0.9111
2号轴承	0.9235
转子	0.9675

由于  $T_{P1} = 5900h, T_{P2} = 14350h, T_{P2}$  不是  $T_{P1}$  的倍数。而为了更换转子, 由表 45.5-17 可知, 必须拆卸一个轴承和轴套。显然, 同时更换它们比一个一个地分开更换更为经济。最好能使  $T_{P1}$  与  $T_{P2}$  互成倍数。如图 45.5-31 所示。

由图 45.5-31 可知, 最佳的  $T_{P2} = 12000h, T_{P1} = 6000h$ , 虽然第 1 组的维修费从 0.08050 元/h 上升为 0.08077 元/h (图 45.5-29), 第 2 组的维修费从 0.01138 元/h 上升为 0.01262 元/h (图 45.5-30), 但系统的维修总费用下降了。

第一种情况: 如果取  $T_{P1} = 5900h, T_{P2} = 14350h$  时根据  $T_{P1}$  和  $T_{P2}$  进行预防维修, 则维修总费用为 2 次按  $T_{P1}$ 、1 次按  $T_{P2}$  进行预防维修的费用的和

$$C_T = (2 \times 0.08050 + 0.01138) \text{ 元/h}$$

$$= 0.17238 \text{ 元/h}$$

第二种情况:如果取  $T_{P1} = 6000\text{h}$ ,  $T_{P2} = 12000\text{h}$ , 则  
第 1 次按  $T_{P1}$  进行的维修费 = 0.08077 元/h

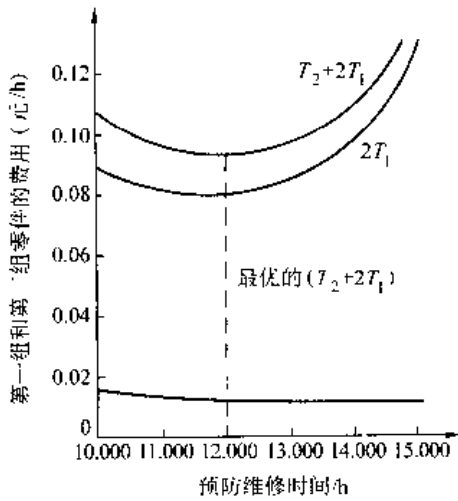


图 45.5-31 第一组、第二组零件的维修总费用与预防维修时间的关系

按  $T_{P2}$  进行的维修费,包括工人工资、停工时间 = 0.01262 元/h

轴承更换费 = 0.00916 元/h

轴套更换费 = 0.00434 元/h

第 2 次按  $T_{P1}$  进行的维修费 = 0.04990 元/h

所以,第二种情况的总费用为

$$C_T = 0.15679 \text{ 元/h}$$

可见,第二种情况的维修总费用较低。

由于  $T_{P1}$  增大,所以第 1 组零件的可靠度下降,而由于  $T_{P2}$  减小,所以第 2 组零件的可靠度增大,如图 45.6-28 所示。

第 1 组、第 2 组零件的可靠度如表 45.5-22。

表 45.5-22 同意接受的可靠度

零件	可靠度
轴	0.9792
轴套	0.9598
1号轴承	0.9096
2号轴承	0.9225
转子	0.9814

综上所述,本例说明了维修费用怎样能够在可以接受的可靠性水平下达到最少,同时,用户不应仅仅根据基本费用来优化系统的可靠性,必须也考虑最佳维修周期。

### 7 可靠性、维修性费用的计算模型

可靠性及维修性学科的应用对降低全寿命费用起着关键的作用,因为可靠性决定着故障的频度,而维修性决定着故障的修理时间。系统设计师必须权衡性能、

可靠性、维修性及生产指标,以便减少寿命期费用。

为满足这一要求,应建立根据某一全寿命费用模型导出的一种权衡的设计方法,该全寿命费用模型由计算可靠性、维修性及费用变量的子模型组成,并由这些子模型所决定。全寿命费用模型包括描述采购费用及维修与后勤保障费用的子模型,这些子模型受功能目标及最低性能要求的限制,见图 45.5-32。

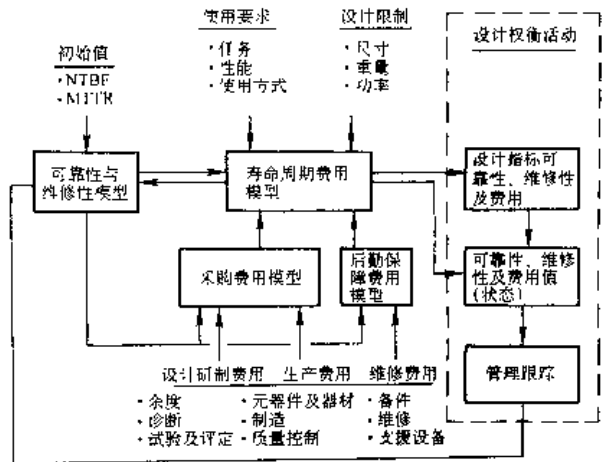


图 45.5-32 可靠性、维修性与费用的设计模型

为了研究降低全寿命费用,需要推导出系统固有有效度费用,由式 (45.5-5) 有

$$A(t) = \frac{MTBF}{MTBF + MTTR} \tag{a}$$

式 (a) 可写为

$$\alpha = \frac{MTTR}{MTBF} = \frac{1-A}{A} \tag{b}$$

对规定的有效度  $A$  来说,  $MTTR$  及  $MTBF$  必须是处于某一恒定比值,即提高其中的一个参数时,也要提高另一个参数,以使所要求的有效度值保持恒定。如果假设每个  $MTTR$  值有一个具体费用,而每个  $MTBF$  同样也要有一个费用,则

$$C_A = C_M - C_R = F(MTTR) + g(MTBF) \tag{45.5-90}$$

式中  $C_A$  ——有效度费用;

$C_M$  ——维修性费用;

$C_R$  ——可靠性费用。

希望使  $C_A$  最小(获得满足规定有效度的最低费用系统)并满足式 (b) 的条件。但是,对于某一给定的  $MTTR$  及  $MTBF$  来说,  $C_M$  及  $C_R$  本身是可变的。可能有许多方法获得某一具体的  $MTBF$ , 每个  $MTBF$  有一个最低费用值的获得方法, 每个  $MTTR$  也是如此。

改变可靠度  $R$  (通过设备设计获得) 及维修度  $M$  (通过测试设备、修理系统及备件供给原则获得), 以便得出规定的(任务要求的)有效度  $A$ , 三者之间的关系

见式 (a)。应注意, MTBF 是可靠性的度量, MTTR 是维修性倒数 (1/M) 的度量。可靠性、维修性与有效度之间的关系, 如图 45.5-33 所示。有效度是一空间曲面。图中的  $A_1, A_2, A_3, A_4$  为等有效度曲线, 从  $A_1$  到  $A_4$  逐个升高。这些线不表示费用。

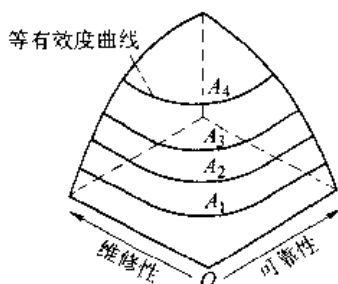


图 45.5-33 A-R-M 关系曲面

对于费用来讲, 仍可用图 45.5-33 所示的曲面原理。但  $A, M$  和  $R$  则表示有效度费用、维修度费用和可靠度费用。由于 MTTR 与  $1/M$  成正比, MTBF 与  $R$  成正比, 故式 (b) 可写为

$$\frac{1}{MR} = \left( \frac{1-A}{A} \right) \quad (45.5-91)$$

或

$$M = \left( \frac{1-A}{A} \right) \frac{1}{R} \quad (45.5-91)$$

上式取偏导数, 有

$$\frac{\partial M}{\partial R} = \left( \frac{1-A}{A} \right) \left( -\frac{1}{R^2} \right) \quad (45.5-92)$$

因此, 对任一有效度来说, 当可靠度提高时, 维修度  $M$  与可靠度的  $R^2$  成比例下降。

如对图 45.5-33 作  $R-M$  平面的投影, 则可得图 45.5-34 所示的有效度  $A$  与  $R$  及  $M$  相关曲面的两维投影图, 图中的曲线  $A_1, A_2, \dots$  表示给定有效度  $A_i$  时, 可以由许多不同的可靠性值及维修性值组合得到。

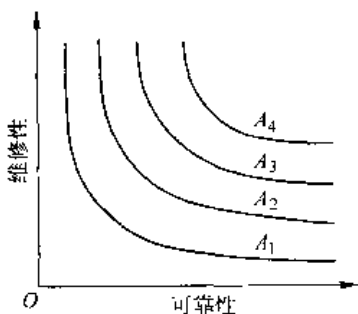


图 45.5-34 A-R-M 曲面的两维投影图

图 45.5-35 给出一系列的“等费用”曲线, 一条曲线代表某一具体预算可能采取的可靠性及维修性的组合。预算从  $B_1$  增加到  $B_4$ , 这些曲线与坐标轴的交点表示某一具体的预算只花在可靠性 (或维修性) 上, 而不花在其他方面。等费用线不表示有效度, 在某一具体的等费用线上可能有许多不同的有效度值。例如图

45.5-35 中任一等费用线上的一点  $C$ , 表示某一固定预算 (在这种情况下是  $B_1$ ) 可能选取的维修性费用及可靠性费用值 ( $qM, qR$ )。当所有费用都花在维修性时, 维修性的单位费用为  $B_1/qM$ , 这里  $qM$  为  $C$  点的纵坐标,  $qM_1$  为  $B_1$  与纵轴的交点; 同理可得, 当所有费用都花在可靠性时, 可靠性的单位费用为  $B_1/qR$ ,  $qR_1$  为  $B_1$  与横轴的交点。

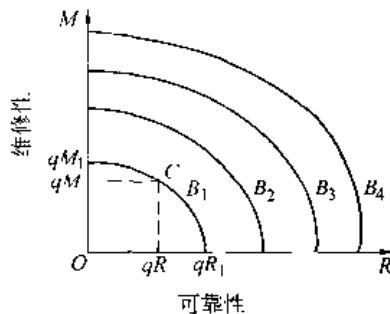


图 45.5-35 等费用曲线

图 45.5-36 是将图 45.5-34 中的等有效度曲线与图 45.5-35 中的等费用曲线结合起来。使每一  $B$  曲线与每一  $A$  曲线相切, 得切点  $P_1, P_2, \dots$ 。这些切点, 为满足每一有效度要求的最低费用的可靠性及维修性组合。

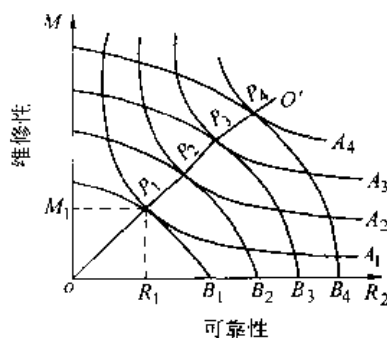


图 45.5-36 维修性与可靠性的最佳结合

应指出, 任一个切点, 两条曲线的斜率必须相等。因此, 对某一给定的有效度曲线, 在最优的情况下,  $R/M$  (可靠性及维修性之比) 的斜率必须等于  $C_M/C_R$  (预算曲线的斜率)

由  $P_1$  到  $P_4$  的连接曲线  $OQ$ , 是以最低费用获得给定有效度下的可靠性及维修性的最优组合曲线, 表示在每种情况下, 为了以最低费用增加  $A$  必须增加的  $R$  及  $M$ 。这是为获得  $A$  而在  $R$  及  $M$  之间权衡的经济手段。

在设计中需要直观地找出最佳点或区段, 为此, 先作图 45.5-37, 图中曲线  $A_1$  为等有效度曲线, 与各等费用曲线分别相交于  $C_1, C_2, \dots, C_4$  各点 (费用下降), 再交于  $C_5, C_6, \dots$  (费用上升), 这样, 就得到了各可靠度  $R$  下的费用。



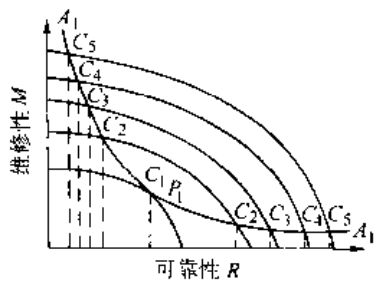


图 45.5-37 等有效度曲线上的费用

将图 45.5-37 中曲线  $A_1$  与等费用曲线各交点的数据, 横轴仍为可靠度, 纵轴为费用描点, 画得不同可靠度  $R$  的费用曲线。见图 45.5-38。费用最低的点  $P_1$  是获得固定的有效度的最低费用点。

此外还可以用相似的方法来解决以某固定的费用得到最高的有效度的问题。选择某一固定的等费用曲线, 并找到相切的有效度曲线。任何更高的有效度曲线

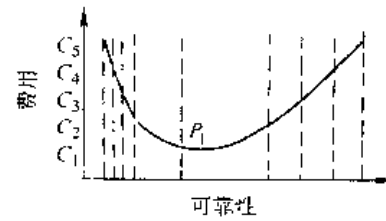


图 45.5-38 U形费用曲线

在任何点上不与等费用曲线相交, 不可能用规定的预算获得更高的有效度, 任何比相切曲线低的有效度曲线表示比在规定的预算条件下的最大有效度值低的有效度值。应指出, 在某一固定费用下获得最大的有效度的问题和在某一固定的有效度下使费用最低的问题在形式上是相同的。在理论上, 从哪一方面着手都可以, 而且会得到同样的结果。

## 第 6 章 可靠性分析中的统计分析软件应用

### 1 概述

20 世纪以来,由于电子计算机应用技术的高速发展,使人类从各种繁重的体力和脑力劳动中得到解放,如今人们可以借助电子计算机及其统计分析系统或统计软件包来实现繁冗的数据处理工作,在大量的信息中获得有科学价值的结果,帮助人们分析、决策和解决问题。用电子计算机实现统计数据处理,必须对采集的信息进行存储、整理、计算、统计分析和图表绘制,从而利用计算的结果揭示和阐明客观事物变化规律。

目前,经典的大型全功能的统计分析系统或统计软件包,包括 SAS, SPSS, SYSTAT, STATISTICA, BMDP, SPLM, MiniTab 等。

美国俄克拉荷马州 (Oklahoma) Stat Soft 软件公司继 1993 年推出 Windows 版本大型专业统计图表分析软件包 STATISTICA for Windows (v4.5) 之后,近年来又推出 v6.0 的 STATISTICA for Windows 版本,它除保持了原有版本的诸多优点外,还汇集了 SPSS, SAS, SYSTAT 与 BMDP for Windows 的长处,增加了更多的统计分析 with 图表分析功能,可以说是当今国际上最新的 Windows 版本统计分析 with 图表分析软件包之一。它的主要特点是:

1) STATISTICA/Win 不但可以在 Windows95/98/2000/ME/XP 下运行,而且可在 Windows NT 的各种版本下运行。

2) 运行速度比 SPSS/Win, SAS/Win, SYSTAT/Win 以及 BMDP/Win 更快。

3) 菜单栏上详细的统计专业名词,方便用户易于理解其功能,除帮助功能外,在对话框上都有简明扼要的解释,并可配合中文使用,增加可读性。

4) 每一个统计或图形模块均能生成图标(按钮),而且图标形象直观,便于判别和理解,用户可以根据不同的需要将自己常用的统计或图形模块生成用户自动任务按钮(Customize Auto Task Button),减少利用菜单选择的时间。

5) 有功能强大的数据管理系统,可同时打开多个数据文件,而且还可以在分析的过程中打开数据文件。

6) 可分步进行分析,用户可根据需要选择所需的统计或图表模块。

7) 作图(制作图表并进行分析)速度比 SPSS/Win, SAS/Win 更快,且能制作更多类型的统计图表,

对于图形的颜色,线条的类型有更多的选择,为用户制作多种类型的统计图表提供良好的环境,还具有用户自定义统计图表的功能,用户可以根据需要定义具有自己特色的统计图表。

8) 输出结果简单明了,对于具有显著性意义( $P < 0.05$ )的结果以不同颜色(例如红色)表示,输出结果可以用指定文件或标准文本文件格式存盘,便于资料的再开发,还可以把图表、数据计算结果编辑成图文并茂的报告。

#### 1.1 运行环境

STATISTICA/Win v6.0 能在 Windows95/98/2000/ME/XP 或 Windows NT3.51 以及更高版本下进行运算,它推荐的基本设置要求如下。

1) 一个能运行 Windows95/98/2000 操作系统的 PC 机,本手册的实例是在 Windows98 操作系统支持下运行的。

2) 至少需要 16MB 内存。

3) S-VGA 显示器以及装有与 Windows98/2000 相兼容的高级图形适配卡。

4) 整个 STATISTICA/Win 约占据 20MB 硬盘空间。

5) 双速或四速以上 CD-ROM 驱动器及至少有一个 3.5in 的软驱。

6) 支持 Windows98 或 Windows2000 的图形打印机一台。

7) 鼠标(Mouse)。

#### 1.2 基本功能

STATISTICA/Win 统计分析有 15 个主命令,130 多个子命令。主要功能包括:

基本分析/制表 (Basic Statistics/Table);

非参数统计检验/分布拟合 (Nonparametrics/Distrib);

广义单因素方差分析与多因素方差分析 (General ANOVA/MANOVA);

多重回归分析 (Multiple Regression);

非线性估计法 (Nonlinear Estimation);

时间序列/预测预报分析 (Time Series/Forecasting);

聚类分析 (Cluster Analysis);

因子分析 (Factor Analysis);  
 典型分析 (Canonical Analysis);  
 多维尺度分析 (Multidimensional Scaling);  
 路径分析 (SEPATH);  
 信度/项目分析 (Reliability/Item Analysis);  
 判别分析 (Discriminant Analysis);  
 对数线性模型分析 (Log-linear analysis);  
 生存分析 (Survival Analysis)。

### 1.3 启动与退出

#### 1.3.1 启动

STATISTICA/Win 的启动可以按如下 3 种方法进行。

方法一 开机后,启动 windows95/98/2000,在桌面上如图 45-6-1 所示的快捷图标处双击鼠标左键,立即显示 STATISTICA/Win 的 STATISTICA Module Switcher (STATISTICA 模型切换)对话框(见图 45-6-2)。

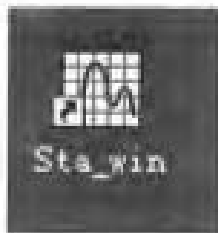


图 45-6-1 STATISTICA/Win 图标

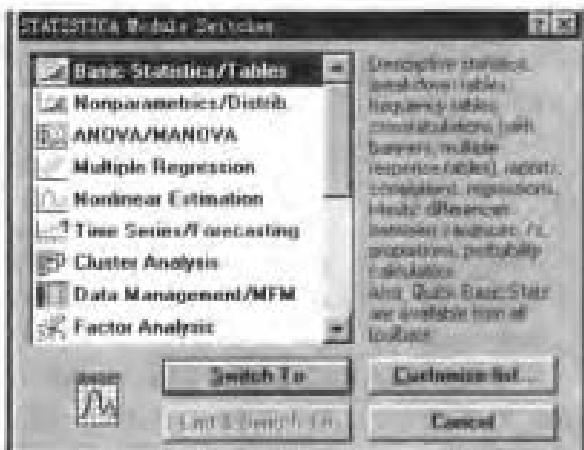


图 45-6-2 STATISTICA Module Switcher (STATISTICA 模型切换)对话框

方法二 开机后,启动 Windows95/98/2000,单击 [开始] ⇒ [程序] ⇒ (双击) [STATISTICA] ⇒ [STATISTICA],也立即显示 STATISTICA/Win 的 STATISTICA Module Switcher 对话框(见图 45-6-2)。

方法三 开机后,启动 Windows95/98/2000,双击桌面上“我的电脑”快捷图标(图 45-6-3),然后选用地址如 D:\ → Stat,最后,双击“StatSoft 软件”图标(见图 45-6-4),也会显示 STATISTICA/Win 的 STATISTICA Module Switcher 对话框(见图 45-6-2)。

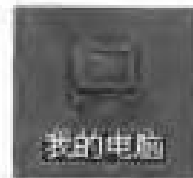


图 45-6-3 “我的电脑”图标



图 45-6-4 StatSoft 图标




STATISTICA Module Switcher (STATISTICA 模型切换)对话框(见图 45-6-2)有 3 个区:

左侧(第一区)提供了本统计图表软件可选用的模型(包含数据管理、统计分析与图表制作等模块);

右侧(第二区)显示当前模块的必要功能说明;

下部(第三区)有 4 个选项: Switch To (切换), Customize list (显示常用模型清单), End & Switch To (结束并切换), Cancel (删除当前对话框)。

#### 1.3.2 退出

当进入 STATISTICA/Win 的任何对话框,其右上角都有图标 。用户如果想终止操作,可以单击 ,立即关闭该对话框,往下继续单击 ,最终可以退出 STATISTICA/Win,返回到 Windows 桌面。

## 2 数据文件的建立

### 2.1 统计数据类型

统计分析 (Analysis) / 图表分析 (Graphs/Tables) 中遇到的数据一般可分为 6 类:计量数据 (Measurement Data), 计数数据 (Enumeration Data), 等级数据 (Ranked Data), 有序数据 (Ordinal Data), 名义数据 (Nominal Data) 与删失数据 (Censored Data)。各种类型的数据又可据实际分析需要而相互转化。一般说来,不同类型的数应采用不同的统计/图表分析方法。在机械可靠性分析中主要采用下列三

种数据类型:

(1) 计量数据 (Measurement Data) 或称数值数据 (Numerical Data)

例如, 测定疲劳寿命、材料的强度、零件所受的压力等, 这一类数据一般具有计量单位, 原则上它的取值可以在某一区间范围内的任一实数, 通常称这类数据是连续数据 (Continuous Data), 它的统计分析与其具有密度的连续随机变量 (Continuous Random Variable) 的分布有关, 在一般的统计分析软件包中, 这是主要的、大量的、有时甚至是全部的分析对象。

统计分析时常用均数、标准差、 $t$  检验、单因素或多因素方差分析、相关或回归分析等。

(2) 计数数据 (Enumeration Data) 或称分类数据 (Categorical Data)

例如, 某一疲劳寿命相同的零件个数, 某材料温度在不同范围内的频数等, 这一类数据一般是可数的, 它们的取值范围是整数 (个数), 通常情况下, 是取非负整数, 称这类数据是计数的, 或称被考察的指标是计数的。它的统计分析是与具有离散随机变量 (Discrete Random Variable) 的分布有关, 在一般的统计分析软件包中, 由于实际工作的需要, 目前也成为要分析的对象。

如统计分析时常用: 率、构成比、 $\chi^2$  检验、非参数统计检验等。

(3) 删失数据 (Censored Data), 或称截尾反应数据 (Censored Responses), 不完全数据 (Un-complete Data)

例如, 在疲劳领域中, 疲劳实验研究的一部分试样可以观察到失效, 能得到准确的疲劳寿命, 这一类数据称为完全数据 (Complete Data), 或非截尾反应数据 (Un-censored Responses), 但往往有一部分试样由于各种原因 (如经费、时间、必要性等), 不能观察到其真正的寿命值, 那有起始时间, 没有确切的终止时间, 但能得到“该试样的寿命不小于某个数值”这样一个信息, 习惯上, 在该数据右上角标以 “+” 号表示。

统计分析时常用: 中位数、生存率、生存分析等。

实际选用 STATISTICA/Win 时, 用户应根据统计设计的目的与要求, 针对不同类型的数据特征, 选用恰到好处的统计/图表分析模块进行分析, 才可能得到正确的结果。

## 2.2 STATISTICA/Win 的文件类型

STATISTICA/Win 自身的文件类型如下:

数据文件 扩展名是: \*.sta, 其数据文件内容可

调入 STATISTICA/Win 使用。

统计分析结果文件 扩展名是: \*.scr, scroll sheet (滚动工作表) 文件, 或扩展名是: \*.rif, 输出结果文件。

图表结果文件 扩展名是: \*.stg, 或扩展名是: \*.whf, Metafile (Windows 应用程序的通用格式文件, 或苹果文件, 或 Meta 文件), 扩展名是: \*.bmp, bitmap (位图)。

## 2.3 建立数据文件

### 2.3.1 数据/大数据文件管理

STATISTICA/Win 除了有众多的统计分析 with 图表分析功能以外, 还有十分完善的数据管理功能。

1) 启动 STATISTICA/Win 以后, 会显示 STATISTICA/Win 的 STATISTICA Module Switcher (STATISTICA 模型切换) 对话框 (见图 45-6-2)。

2) 单击 [Data Management/MPM] (数据管理与大数据文件管理器, Data Management and Megafile Manage), 如图 45-6-5 所示的对话框。

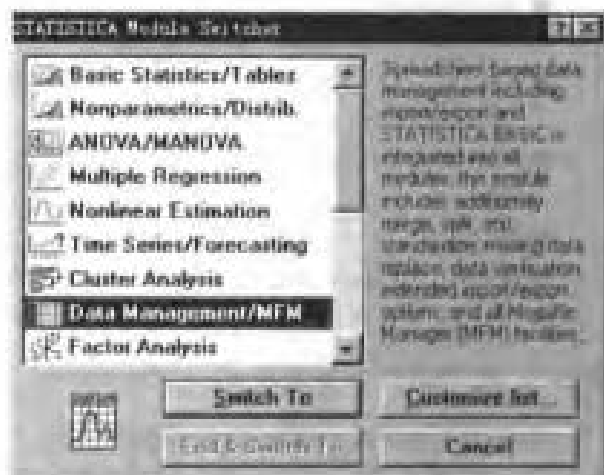


图 45-6-5 STATISTICA Module Switcher (STATISTICA 模型切换) 对话框

3) STATISTICA 模型切换器对话框 (见图 45-6-5) 的右侧, 提供“数据管理与大数据文件管理器”主命令功能简明扼要的讲解, 再单击 [Switch To] (切换到), 得到如图 15-6-6 所示的对话框, 图中数据管理器 (Data Management) 有 13 个子命令。

● Create new data file, 建立新的数据文件。

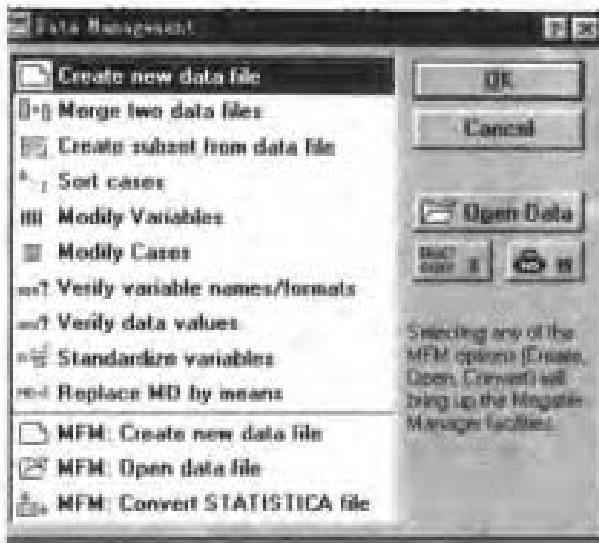


图 45-6-6 Data Management (数据管理) 对话框

- Merge two data file, 合并两个数据文件。
- Create subset from data file, 从已知的数据文件中建立一个子数据文件。
- Sort cases, 个案排序。
- Modify Variables, 修改变量。
- Modify Cases, 修改个案。
- Verify variable names/formats, 校验变量名或格式。
- Verify data value, 校验数据值。
- Standardize variables, 标准化变量。
- Replace MD by means, 用均数替代缺失数据 (Missing Data)。
- MFM: Create new data file, 建立新的大数据文件。
- MFM: Open data file, 打开大数据文件。
- MFM: Convert STATISTICA file, 转换 STATISTICA 文件。

### 2.3.2 建立新的数据文件

利用 STATISTICA/Win 软件对数据进行分析, 首先要建立数据文件, 下面用一个实例 (表 45-6-1) 说明数据的录入方法, 即建立新的数据文件过程。

表 45-6-1 某实验记录资料

编号 (no)	实验员姓名 (name)	实验日期 (date)	温度 T/°C	实验结果 (result)
1	李敏	2003.09.18	22.21	1500
2	叶静	2003.09.24	20.53	1150
3	李敏	2003.10.15	15.69	1200
4	叶静	2003.10.16	16.84	1345
5	李敏	2003.10.17	14.34	1125

#### (1) 启动 STATISTICA/Win

单击 [Data Management/MFM] ⇒ [Switch To] ⇒ [Create new data file] (建立新数据文件) ⇒ [OK], 得到如图 45-6-7 所示的对话框。

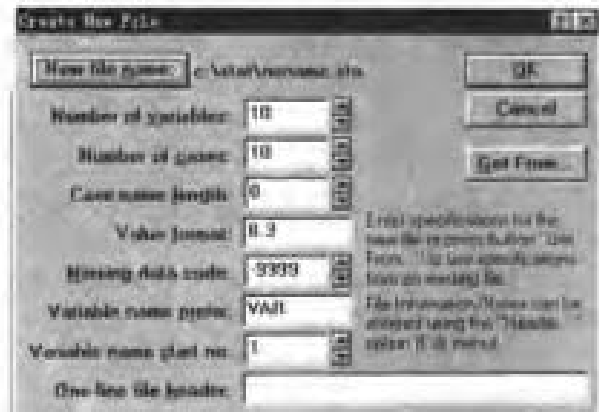


图 45-6-7 Create New File (建立新数据文件) 对话框

Create New File (建立新数据文件) 对话框 (图 45-6-7) 的右下方是本对话框的说明。

新数据文件名 (New file name), 本例选用 C:\statdata\new12.sta, 即新数据文件 new12.sta 存放在硬盘 C 的子目录 statdata 中。图 45-6-7 对话框中有 8 个选项。

- Number of variables, 变量个数, 本例是 5 个 (默认是 10 个)。
  - Number of cases, 个案数, 本例是 5 个 (默认是 10 个)。
  - Case name length, 个案名长度, 默认是 0。
  - Value format, 数值格式, 数值型的宽度默认是 8.3, 即宽度是 8 位, 小数是 3 位, 本例小数是 2 位。
  - Missing data code, 缺失数据代码, 默认是 9999。
  - Variable name prefix, 变量名的前缀, 默认是 VAR。
  - Variable name start no, 变量名的起始号码, 默认是 1。
  - One-line file header, 在线文件抬头, 默认是空白, 此外, 也可以从外部获得 Get From (数据文件)。
- (2) 建立新的数据文件格式

1) 单击 [New file name] (新数据文件名), 选用 d:\statdata\new12.sta。

2) 本例的变量个数 (Number of variables) 是 5 个, 将 10 改为 5。

3) 本例数值的格式是 8.2, 其余选用默认格式, 单击 [OK], 得到如图 45-6-8 所示的对话框。

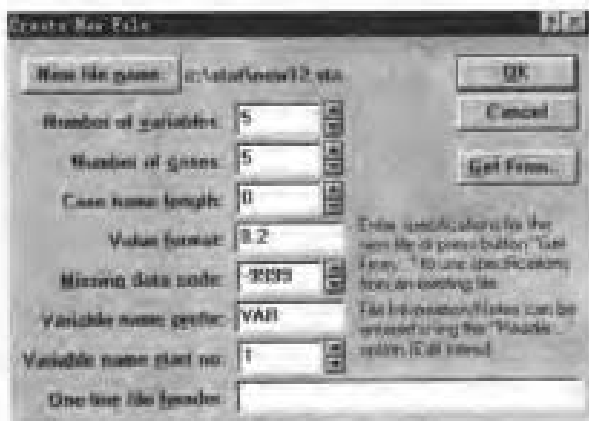


图 45-6-8 建立新数据文件: new12.sta 格式

4) 在图 45-6-8 中,再单击 [OK],得到如图 45-6-9 所示的工作单。



图 45-6-9 new12.sta 工作单

new12.sta 工作单(图 45-6-9)的第 1 行左上角表明当前的数据文件是 new1.sta,它有 5 个变量,5 个个案(5v = 5c)。如果没有特殊要求,可以立即输入数据。由于本例有一些具体的要求,因此必须进一步设

置各个参数(即建立表头)。

5) 在工作单中双击 [VAR1],得到如图 45-6-10 所示的对话框。

本例的第 1 个变量(Variable 1)是 NO,中文标识是“编号”,所以在变量名(Name)中键入 NO,在缺失数据代码(MD code)中取默认值-9999。

在显示格式(Display Format)中,列宽度(Column width)选用 4,而小数(Decimals)选用 0,变量的种类(Category)有数值型(Number),日期型(Date),时间型(Time),科学型(Scientific),货币型(Currency)与百分数型(Percentage)等 6 种。本例编号选用数值型。

描述数据(Representation)选用默认格式。

在长变量名(Long name)(标识,连续,或函数的公式)(label, link, or formula with Functions)中,键入“编号”,然后单击 [OK],得到第 1 个变量名 NO。

6) 双击 [VAR2],得到如图 45-6-11 所示的对话框。

第 2 个变量(Variable 2)是 NAME,列宽度是 8,小数是 0,变量类型是货币型,长变量名是实验员姓名,其余选用默认格式。

7) 单击 [OK],再双击 [VAR3],得到如图 45-6-12 所示的对话框。

第 3 个变量(Variable 3)是 date,列宽度是 10,小数是 0,变量类型是日期型,描述是 03/06/1987,长变量名是实验日期,其余选用默认格式。

8) 单击 [OK],再双击 [VAR4],得到如图 45-6-13 所示的对话框。

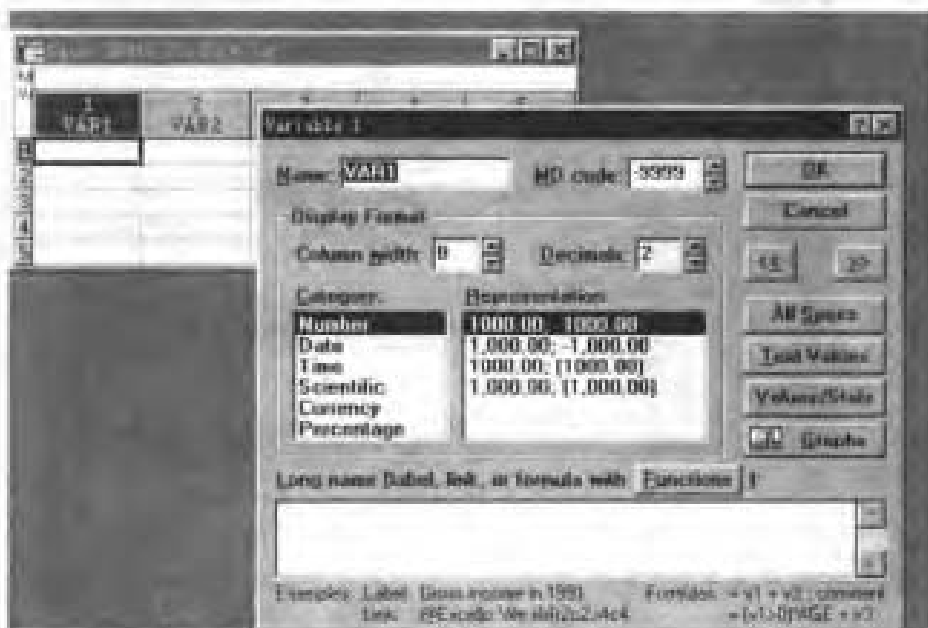


图 45-6-10 Variable 1 (变量 1) 对话框

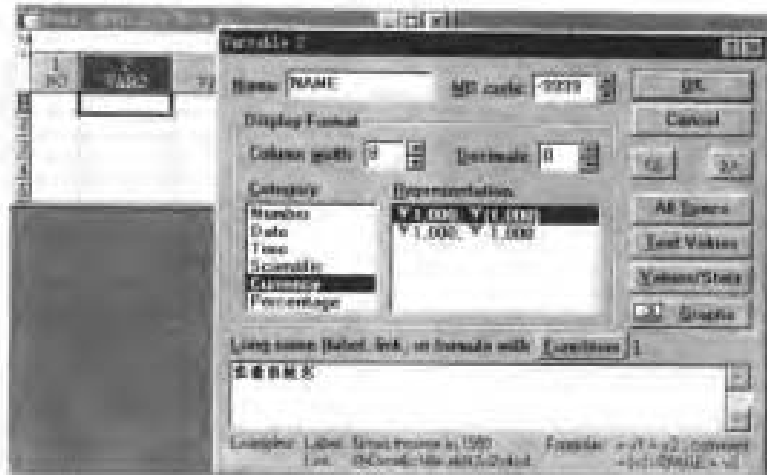


图 45-6-11 Variable 2 (变量 2) 对话框

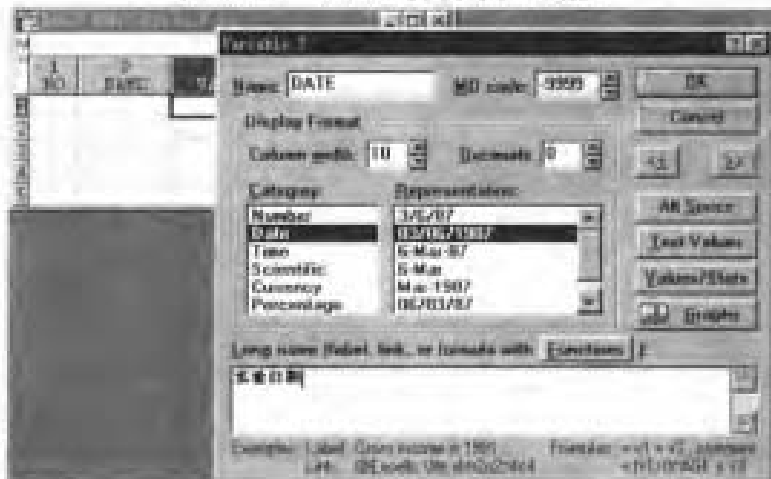


图 45-6-12 Variable 3 (变量 3) 对话框

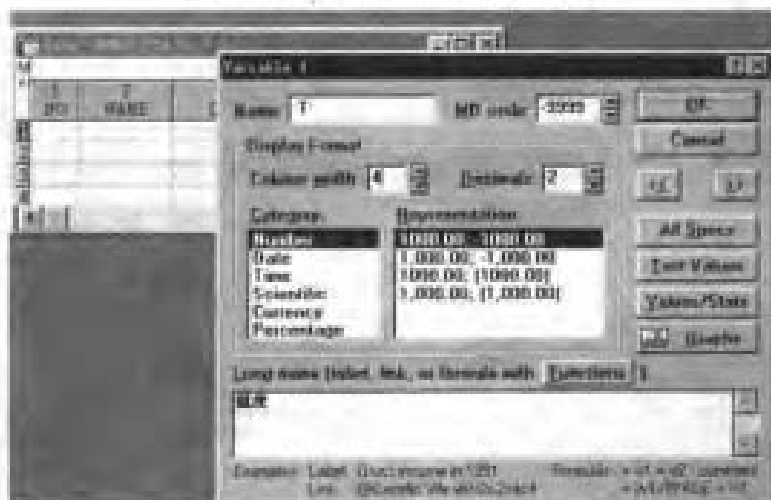


图 45-6-13 Variable 4 (变量 4) 对话框

第 4 个变量 (Variable 4) 是 T, 列宽度是 4, 小数是 2, 变量类型是数值型, 长变量名是温度, 其余选用默认格式。

9) 单击 [OK], 再双击 [VAR5], 得到如图 45-6-

14 所示的对话框。

第 5 个变量 (Variable 5) 是 RESULT, 列宽度是 5, 小数是 2, 变量类型是数值型, 长变量名是结果, 其余选用默认格式。

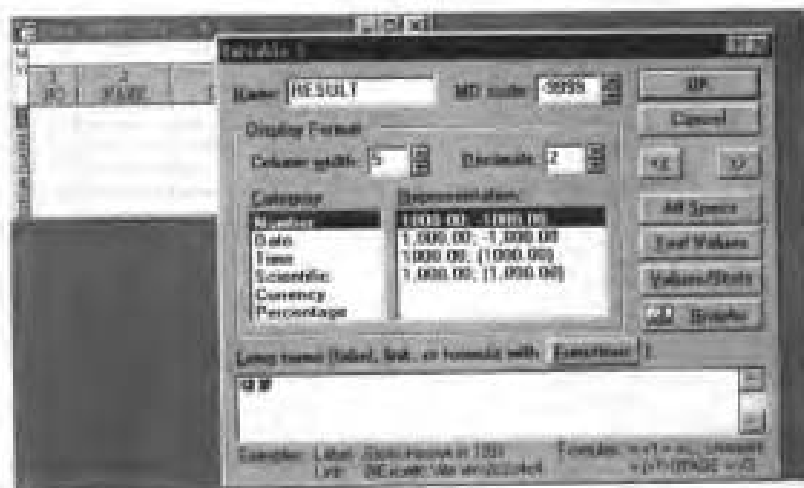


图 45-6-14 Variable 5 (变量 5) 对话框

10) 单击 [OK], 得到 new12.sta 工作单 (见图 45-6-15)。

1	2	3	4	5
NO	NAME	DATE	T	RESULT

图 45-6-15 new12.sta 工作单

11) 依次键入数据, 便得到数据文件 new12.sta (见图 45-6-16), 可将此数据文件保存为: d:\stadata\new12.Sta, 以备后用。

1	2	3	4	5
NO	NAME	DATE	T	RESULT
1	李静	09/18/2003	27	2.1500
2	叶静	09/18/2003	20	5.1150
3	李静	10/15/2003	15	7.1200
4	叶静	10/16/2003	10	2.1345
5	李静	10/17/2003	10	2.1175

图 45-6-16 new12.sta 数据文件

### 2.3.3 数据文件个案排序

STATISTICA/Win 的个案排序 (Sort Cases) 功能可将数据文件中的变量 (数值型或字符型) 进行排序, 方法为: 可依据变量数值作为排序关键词由小到大 (升序, Ascending), 或由大到小 (降序, Descending) 排列; 也可依据变量字符, A, B, C, ……; Z 作为排序关键词的顺序 (即字典法) 排序 (升序, Ascending), 或依据变量 (字符), Z, Y, X, A 的顺序 (降序, Descending) 排序。

在可靠性分析中, 可以应用个案排序结果, 通过查序表法, 得到累计频率。

例 试对已知数据文件 hong1.sta 依据变量 (数值) 作为排序关键词由小到大排序。

1) 进入 STATISTICA/Win 以后, 单击 [Data Management/MFM] ⇒ [SwitchTo] ⇒ [Sort cases] (个案排序) ⇒ [Open Date] ⇒ [d:\stadata\hong1.sta], 得到如图 45-6-17 所示的对话框。

1	2	3	4	5	6
NO	T1	T2	A	B	P
25.000	27.000	18.500	8.000	9.000	3.250
25.000	28.000	18.500	8.000	9.700	3.750
27.000	27.000	17.000	8.500	9.300	1.400
21.000	23.000	15.500	7.000	8.700	2.600
27.000	25.500	18.500	7.000	8.700	3.300
27.000	24.000	16.500	8.000	8.500	1.400
27.000	25.500	17.000	9.000	9.100	2.500
27.000	29.000	15.000	8.500	9.900	1.450
24.000	24.000	18.000	7.000	9.300	3.450
24.000	24.500	18.000	9.000	11.000	3.850
24.000	27.000	19.000	8.000	11.800	2.700
29.000	29.500	14.000	8.000	9.400	2.950
26.000	26.000	17.500	8.000	11.000	2.550
29.000	27.000	14.000	7.500	9.900	3.600
24.000	27.000	17.500	8.000	9.900	3.200

图 45-6-17 数据文件: hong1.sta (部分)

2) 再单击 [OK], 得到如图 45-6-18 所示的对话框。

3) 在 Sort Options 对话框中, 排序选择有 7 个关键词供用户选用。而且, 每一个关键词 (Key) 之后, 都可以选择变量 (Var) 或个案名 (Case Name), 升序 (Ascending) 或降序 (Descending), 数值型 (Numeric) 或字符型 (Text values), 当选完前 3 个关键词之后, 如果关键词有 4 个以上, 还可以单击 [More Keys] (选择更多的关键词), 得到如图 45-6-19 所示对话框。



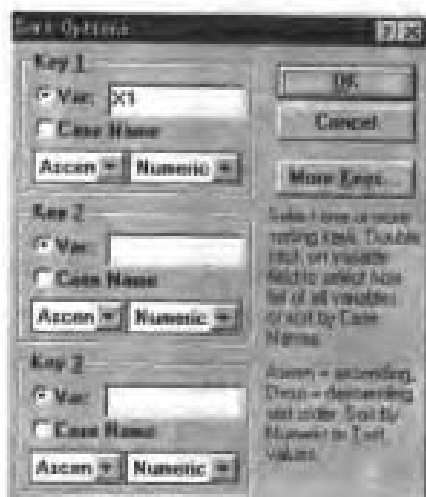


图 45-6-18 Sort Options(排序选择)对话框(一)

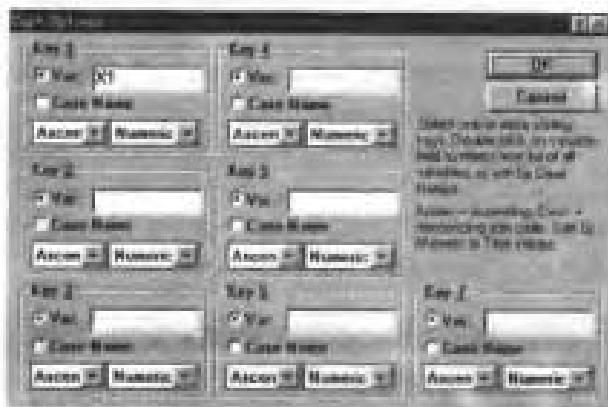


图 45-6-19 Sort Options(排序选择)对话框(二)

4) 单击 [OK], 得到如图 45-6-20 所示的排序后的结果(数据文件)。本结果是排序后的数据文件(部分)(图 45-6-19), 变量 X1 已经从小到大排序。

	X1	X2	X3	X4	X5
21.000	25.000	14.500	8.000	8.700	7.600
22.000	24.000	15.500	8.500	8.550	7.400
22.000	25.500	17.000	9.000	9.100	7.900
23.000	24.000	17.500	8.000	10.400	8.000
23.000	24.000	17.500	7.500	10.700	7.900
22.500	25.000	15.500	8.000	8.700	7.250
22.500	26.500	16.500	7.500	8.700	7.300
23.000	25.000	15.000	8.000	11.800	7.900
23.000	26.000	16.000	8.500	10.700	7.650
23.000	26.500	15.500	8.000	9.100	7.950
23.000	25.000	17.000	8.000	8.300	7.250
23.000	27.000	17.000	8.700	9.300	7.400
23.500	27.000	16.000	8.000	9.900	7.500
23.500	26.000	15.000	7.500	9.800	7.500
24.000	25.000	18.000	8.000	10.200	7.800

图 45-6-20 排序后的数据文件(部分)

本例只对 X1 从小到大排序, 如果同时再选 X2 作为第 2 个关键词(Key2)也是从小到大排序, 其结果是:

X1: 21.0 22.0 22.0 22.0 22.0 22.5 22.5  
23.0 23.0 23.0 23.0 23.0 23.5 23.5...

X2: 23.0 24.0 24.0 24.0 25.0 25.0 25.5  
25.0 25.0 25.0 25.0 27.0 27.0 25.0...

即 X1 从小到大排序, 当 X1 的值相同时, X2 又从小到大排序。类似地, 还可以进行多重(最多 2 重)排序。

### 3 基本统计分析

基本统计分析/制表(Basic Statistics/Tables)包括描述性统计分析(Descriptive Statistics)、相关矩阵分析(Correlation Matrices)、独立样本 t 检验(t-Test for Independent Samples)、相依样本 t 检验(t-Test for dependent samples)、细分与单因素方差分析(Breakdown & One-Way ANOVA)、频数分布分析(Frequency Tables)、制表与标题(Tables and Banners)、概率计算(Probability Calculator)与其他显著性检验(Other Significance Tests)等。

#### 3.1 描述性统计分析

描述性统计分析(Descriptive Statistics)可以输出变量(Variables)的近 20 种统计量(Statistics)、频数分布表(Frequency Tables)、分布情况(Distribution)以及直方图(Histograms)、散点图(Scatter plot)等多种图形(Plots)。

例 某单位实验记录结果如表 45-6-2 所示, 试做描述性统计分析。

表 45-6-2 某单位实验记录结果

74.5	78.8	68.8	78.0	76.4	80.5	80.5	68.7	71.2	73.5
79.5	75.6	75.0	78.8	72.0	72.0	72.0	74.3	71.2	72.0
75.0	73.5	78.8	74.3	75.8	65.0	74.3	71.2	69.7	68.0
73.5	75.0	72.0	64.3	75.8	80.3	69.7	74.3	73.5	73.5
75.8	75.8	68.8	78.5	70.4	71.2	81.2	75.0	76.4	68.0
74.0	72.0	78.5	74.3	78.5	77.6	67.3	72.0	75.0	74.3
73.5	79.5	73.5	74.7	65.0	78.5	81.6	73.4	72.7	72.7
67.2	76.5	72.7	70.4	77.2	68.8	67.3	67.3	67.3	72.7
75.8	73.5	75.0	72.7	73.3	73.5	72.7	81.6	70.3	74.3
73.5	78.5	79.4	76.5	72.7	77.2	64.3	75.0	76.5	70.4

1) 建立数据文件: `descripl.sta (d:\stadata\descripl.sta)`。

2) 选用(单击)[Basic Statistics/Table](基本统计分析/制表), 得到如图 45-6-2 所示的对话框。

3) 再单击[SwitchTo](切换到), 得到如图 45-6-

21 所示的对话框, 选用 [Descriptive statistics] (描述性统计分析)。

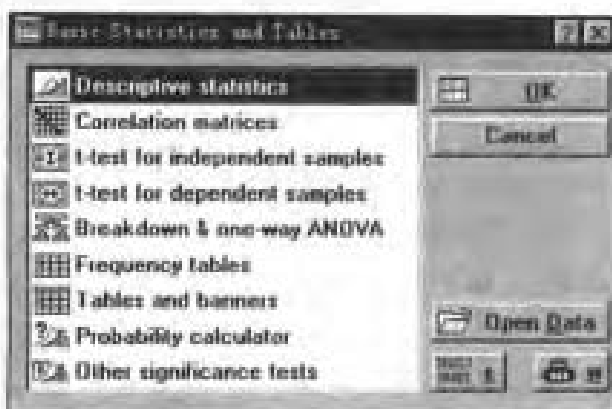


图 45-6-21 Basic Statistics and Tables  
(基本统计与制表) 对话框

4) 在图 45-6-21 中, 单击 [Open Data] (打开数据文件: d:\stadata) ⇒ [OK], 再返回到如图 45-6-21 所示对话框, 单击 [OK], 得到如图 45-6-22 所示的对话框。描述性统计分析 (Descriptive Statistics) 主对话框, 有如下选项。

- Variables, 变量, 分析的变量可以一个或多个, 默认是 none (没有任何变量)。
- Detailed descriptive statistics, 详细列出描述性统计量。
- Options, 选项。

- ☆ Casewise [listwise] deletion of MD, 删除含有缺失数据 (Missing Data) 的个案。
  - ☆ Display long variable names, 显示长的变量名 (含中文标识)。
  - ☆ Extended precision calculations, 扩展计算精确度。
  - Statistics, 统计量。
    - ☆ Median & quartiles, 中位数与四分位数。
    - ☆ Conf. limits for means, 均数的可信限。
    - ☆ Alpha error, 95%, (误差), 默认是 95%, 用户可选用。
    - ☆ More statistics, 更多的统计量。
  - Distribution, 分布分析。
  - Categorization, 分类分析。
- 本例选用的项目与计算:

1) 单击 [Variables] (选用变量), 得到图 45-6-23 所示界面。

在图 45-6-23 所示界面中, 单击 1-SERUM, 得到选择分析的变量 (Select variables) 是 1 (SERUM), 再单击 [OK], 返回到图 45-6-22 所示界面, 这时, 描述性统计分析 (Descriptive Statistics) 的变量是 SERUM。

2) 单击 [Detailed descriptive statistics] (详细列出描述性统计量), 得到如图 45-6-24 所示界面。

整理后得到变量 SERUM (的统计量 (Statistics) (见表 45-6-3)。

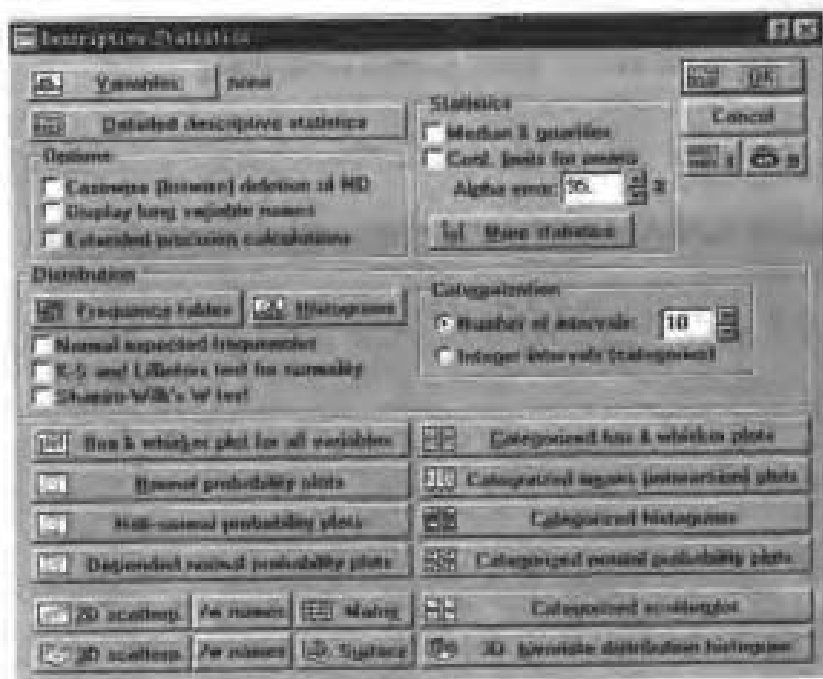


图 45-6-22 Descriptive Statistics (描述性统计分析) 主对话框



图 45-6-23 选择分析变量

Continuous	Missing	Mean	Variance	Minimum	Maximum	Std. Dev.
SERUM	100	73.89500	44.30000	64.30000	84.30000	3.926417

图 45-6-24 描述性统计量结果

表 45-6-3 变量 SERUM 的统计量

	有效例数 Valid N	均值 Mean	最小值 Minimum	最大值 Maximum	标准差 Std. Dev.
SERUM	100	73.89500	64.30000	84.30000	3.926417

### 3.2 独立样本（成组）t-检验

独立样本（成组）t-检验 (T-Test for Independent Samples (Groups)) 又称 Student t 检验（学生氏 t 检验），或成组 t 检验，团体 t 检验，t 检验用以检验两样本分别代表的总体均值是否相等，具体的假设检验依各种问题的不同而异，它以 t 分布为其理论根据。

独立样本 t 检验用以比较两组样本均值差异显著性所需的 Student-t 统计量，给出分组统计量 T 值，并

进行分离方差估计值的 t 检验 (t-Test with Separate Variance Estimates)，多变量霍特林 T<sup>2</sup> 检验 (Multivariate Test (Hotelling's T<sup>2</sup>))，列文方差齐性检验 (Levene's Test (Homogeneity of Variances))，还可以绘制箱形与触须线图 (Box & Whisker Plot)、分类直方图 (Categorized Histograms)、分类正态概率图 (Categorized Normal Prob. Plots)、分类趋势正态图 (Categorized Detrended Normal Plots)、分类散点图 (Categorized Scatter plot) 等。

#### 3.2.1 独立样本 t-检验 (1) (记录/个案格式)

独立样本 t 检验 (t-Test for Independent Sample (Groups)) 可以按一个个案一个记录格式 (分组) 的数据 (One record per case (use a grouping variable)) 进行 t 检验。

例 某材料 1 测得 11 例失效临界值与材料 2 的 13 例失效临界值值如表 45-6-4。问该两种材料的失效临界值是否不同。

1) 建立数据文件，independ.sta (d:\studata\independ.sta)。

2) 单击 [Basic Statistics/Tables] ⇒ [Switch To] ⇒ [t-test for independent samples] (独立样本 t 检验) [Open Data] ⇒ [OK]，得到图 45-6-25 所示界面。

独立样本 (成组) T-检验 (T-Test for Independent Samples (Groups)) 主对话框有如下选项。

表 45-6-4 材料 1 与材料 2 的失效临界值

材料 1	2.69	3.24	3.73	3.73	4.32	4.73	5.18	5.58	5.78	6.40	6.53		
材料 2	1.67	1.98	1.98	2.33	2.34	2.50	3.60	3.73	4.14	4.17	4.57	4.82	5.78

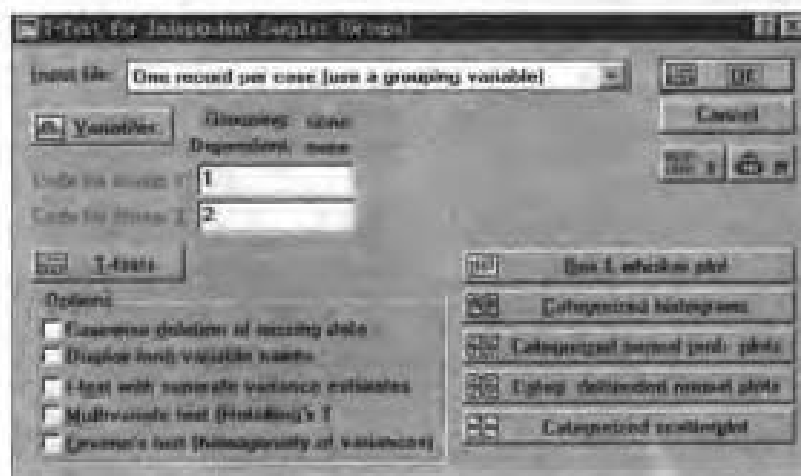


图 45-6-25 T-Test for Independent Samples (Groups) (独立样本 (成组) t-检验) 主对话框

● **Input file**, 输入数据文件, 有两种选择。

☐ 击 **One record per case (use a grouping variable)**, 每一个个案是一个记录 (用分组变量)。此为默认格式。

☐ **Each variable contains the data for one group**, 每一个变量属于一组数据。

● **Variables**, 分析的变量。

☐ **Grouping**, none, 分组变量: 没有任何变量, 默认格式。

☐ **Dependent**, none, 因变量: 没有任何变量, 默认格式。

● **Code for Group1**, 第一组的代码, 默认是 1。

● **Code for Group2**, 第二组的代码, 默认是 2。

● **T-tests**, t 检验, 将进行 t 检验。

● **Options**, 选项如下。

☐ **Casewise deletion of missing data**, 删除全部缺失数据, 默认格式。

☐ **Display long variable names**, 显示长的变量名 (含中文标识)。

☐ **t-test with separate variance estimates**, 有分离方差估计值的 t 检验。

☐ **Multivariate test [Hotelling's T<sup>2</sup>]**, 多变量霍特林 T<sup>2</sup> 检验。

☐ **Levene's test [homogeneity variances]**, 列文方差齐性检验。

● **Box & whisker plot**, 箱形与触须线图。

● **Categorized histograms**, 分类直方图。

● **Categorized normal prob. plots**, 分类正态概率图。

● **Categ. Detrended normal plots**, 分类趋势正态图。

● **Categorized scatter plot**, 分类散点图。

本例选用的项目、计算

1) **Input file**, **One record per case (use a grouping variable)**, 本例数据格式属于此类, 适用此项。

2) 单击 **[Variables]** (分析的变量), 得到图 45-6-25 所示界面。

分组变量 (**Grouping variable**) 是 1-G, 因变量 (**Dependent variables**) 是 2-X, 单击 **[OK]**。

3) 再在选择项 (**Options**) 中作如下选择如图 45-6-27。

4) 单击 **[T-tests]** (进行 t 检验), 得到如下计算结果。

Grouping: G, g=分组 (independent)						
Group1: G=1, 1						
Group2: G=2, 2						
Mean	Mean	t-value	df	T	variance	
G=1, 1	G=2, 2	2.329872 *	22 *	.018668 *	2.539909 *	
	p	Valid N	Valid N	Sig. (2v)	Std. Dev.	
df	2-sided	G=1, 1	G=2, 2	G=1, 1	G=2, 2	
21.35420 *	.019070 *	11 *	13 *	1.502977 *	1.324268 *	
F-ratio	p	Levene	df	F	Levene	
variances	variances	F (1, df)	Levene	Levene		
1.002137 *	1.000000 *	.037990 *	22 *	.547312 *		

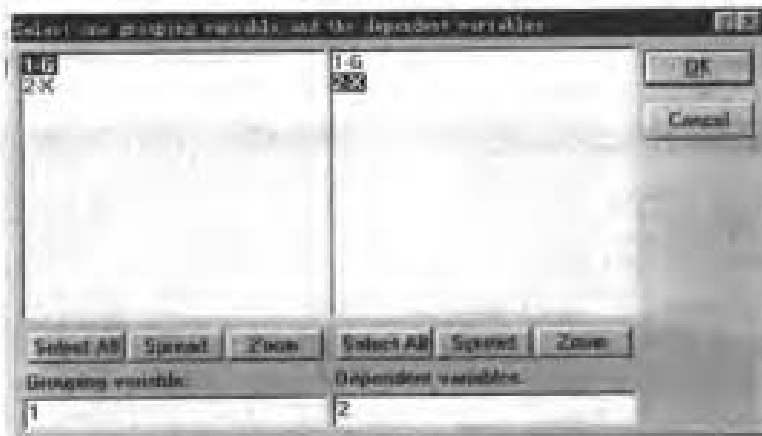


图 45-6-28 选择分组变量与因变量

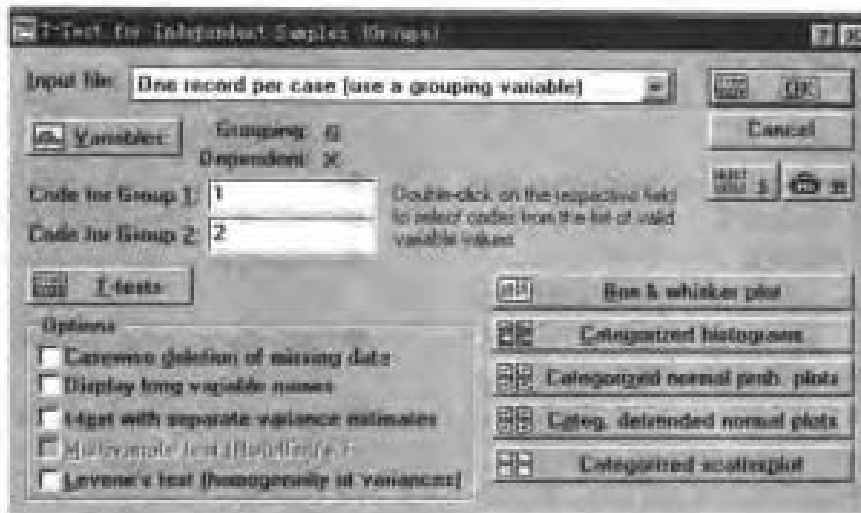


图 45-6-27 T-Test for Independent Samples (Group) (选用后的独立样本 (分组) T 检验) 对话框

### 主要结果分析

得到列文  $p$  值是: 0.847312,  $P > 0.80$ , 可认为方差齐。

两组的均数 (Mean) 分别是: 4.710909, 3.354615,  $t$ -值 ( $t$ -value) = 2.539373,  $P < 0.05$  ( $p = 0.018644$ ), 故可认为材料 1 与材料 2 的失效临界值有显著性差异。

### 3.2.2 相依 (相关) 样本 $t$ 检验

相依 (相关) 样本  $t$  检验 (T-Test for Dependent (Correlated) Samples), 亦称配对  $t$  检验或自身对照, 它对相依或相关两样本均数进行检验并推断其是否有显著性差异。

例 10 例试样经热处理前后的硬度如表 45-6-5 所列, 问经热处理前后的硬度有无提高。

表 45-6-5 热处理前后的硬度值

热处理前	11.3	15.0	15.0	13.5	13.8	10.0	11.0	12.0	13.0	12.3
热处理后	14.0	13.8	14.0	13.5	13.5	12.0	14.7	11.4	13.8	12.0

1) 建立数据文件: paired.sta (d:\stadata\paired.sta) 如图 45-6-28 所示。

2) 选用 (单击) [Basic Statistics/Tables] (基本统

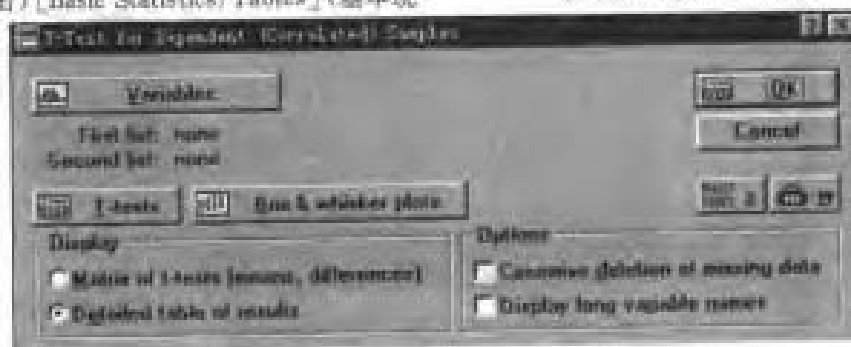
图 45-6-28 数据文件: paired.sta

计分析/制表)  $\rightarrow$  [Switch To]  $\rightarrow$  [T-Test for Dependent (Correlated) Samples] (相依 (相关) 样本  $t$  检验) [Open Data]  $\rightarrow$  [OK], 得到如图 45-6-29 所示对话框, 在对话框中有如下选项:

● Variables, 表示选用的变量。

☆ First list: none, 选用第一组变量清单, 默认是无任何变量 (none)。

☆ Second list: none, 选用第二组变量清单, 默认是无任何变量 (none)。

图 45-6-29 [T-Test for Dependent (Correlated) Samples] (相依 (相关) 样本  $t$  检验) 主对话框

- T-tests, 进行相依(相关)样本t检验。
- Display, 显示。
- ☑ Matrix of t-tests (means, differences), t检验的矩阵(均数, 差异)。
- ☑ Detailed table of results, 结果的详细表, 默认格式。
- Options, 选项。
- ☑ Casewise deletion of missing data, 删除含有缺失数据(missing data)的个案。
- ☑ Display long variable names, 显示长的变量名(含中文标识)。

本例选用的计算与结果

- 1) 单击 [Variables], 选用进行相依(相关)样本t检验的变量, 得到如图 45.6-30 所示对话框。
- 2) 第一组变量清单(First variable list)选用: 1-X1, 第二组(可选)变量清单(Second variable list (optional)) 选用: 2-X2。单击 [OK], 得到图 45.6-31 所示对话框。
- 3) 在选择项(Options)中, 全选, 如图 45.6-31 所示。
- 4) 单击 [T-tests] (进行t检验), 得到计算结果。

STATISTICS						
T-test for Dependent Samples. (paired, etc)						
Marked differences are significant at p<.05000						
Variable	Mean	Std. Dev.	N	Diff.	Std. Dev. Diff.	t
X1	12.59000	1.632615	10	-.680000	1.645725	-1.30662
X2	13.27000	1.080175				
Variable	df	p				
X1						
X2	9	.223742				

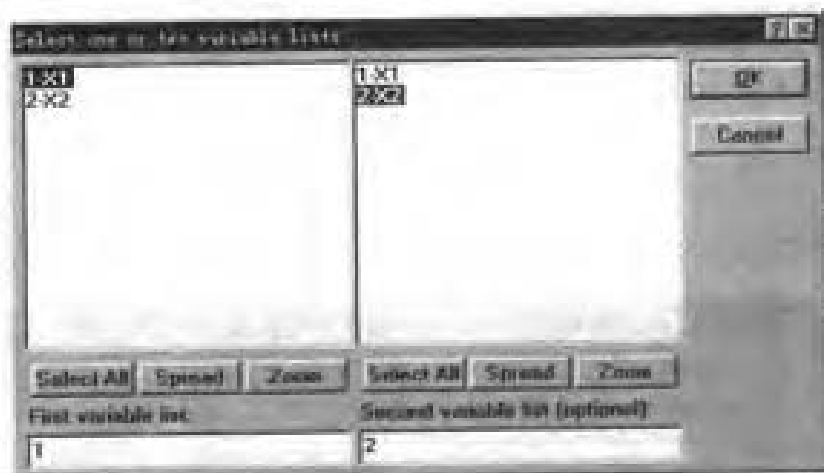


图 45.6-30 Select one or two variable lists (选用一个或两个变量清单)对话框

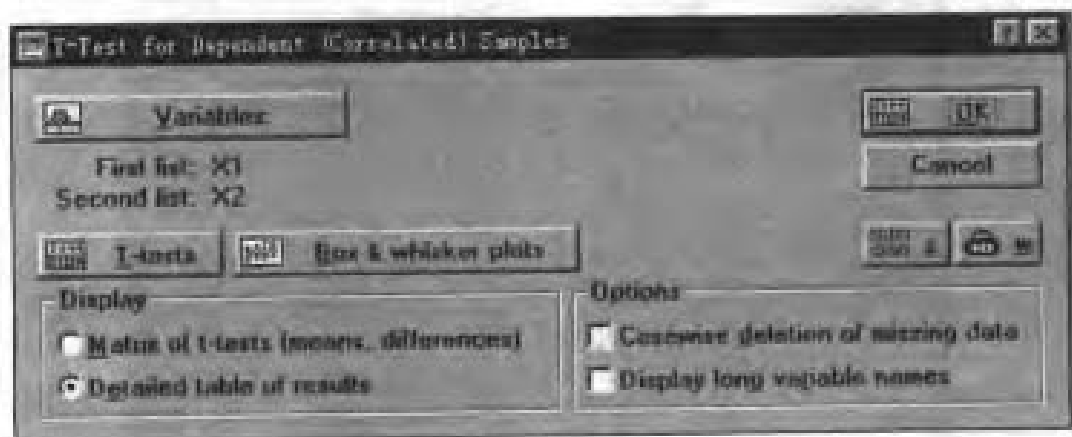


图 45.6-31 已选 T-Test for Dependent (Correlated) Samples (相依(相关)样本t检验)对话框

## 主要计算结果分析

本例有

均数 (Mean) 标准差 (Std. Dev)

热处理前 (X1): 12.59000 1.632619

热处理后 (X2): 13.27000 1.680175

$t = -1.30652$ ,  $P(p = 0.223742) > 0.05$ , 按  $\alpha = 0.05$  水准, 尚不能认为该热处理方法对硬度有作用。

## 4 频数表分析

频数表分析(Frequency Tables)适用于正态分布,除了可以对多个变量进行频数分析以外,还可以进行描述性统计分析(Descriptive Statistics)、变量的正态性检验(Tests for Normality)、绘制直方图(Histograms)、箱形与触须线图(Box & Whisker Plot)、正态概率图(Normal Probability Plots)等多种图形。

其分析步骤为:

1) 建立数据文件。

2) 选用图 45-6-2 主界面,单击【Basic Statistics Tables】(基本统计分析片创表)⇒【Switch TO】⇒【Frequency tables】(频数表分析)⇒【Open Data (打开数据文件)】⇒【OK】,得到图 45-6-32 所示的对话框。

在频数表分析主对话框(见图 45-6-32)中,可选项如下:

● Variables, 频数表分析的变量,可以一个或多个,

默认无任何变量(none)。

● Frequency tables, 频数表分析。

● Histograms, 直方图,对引入分析的变量绘制直方图。

● Descriptive statistics, 描述性统计分析,对引入分析的变量进行描述性统计分析。

● Categorization method for tables & graphs, 频数表与图形的分类方法,可选项:

☆ All distinct values, 所有相异值;默认格式。

☆ with text values, 带文本值;默认格式。

☆ No. of exact intervals, 精确区间个数,默认是 10 个。

☆ "Neat" intervals; app.no., "整齐"区间;应用的个数;默认是 10 个。

☆ Step size 1, 步长,默认是 1。

☆ Starting at or at minimum, 起始;默认是 0,或起始于最小值。

☆ Integer categories, 整数分类。

☆ with text values, 带文本值;默认带文本值。

☆ Specific grouping codes (values), 指定分组变量的代码(值)。

☆ User-specified categories, 用户指定的分类。

● Casewise [listwise] deletion of MD, 缺失数据全部删除。

● Display options, 显示的选择项。

☆ Cumulative frequencies, 累计频数;默认格式。

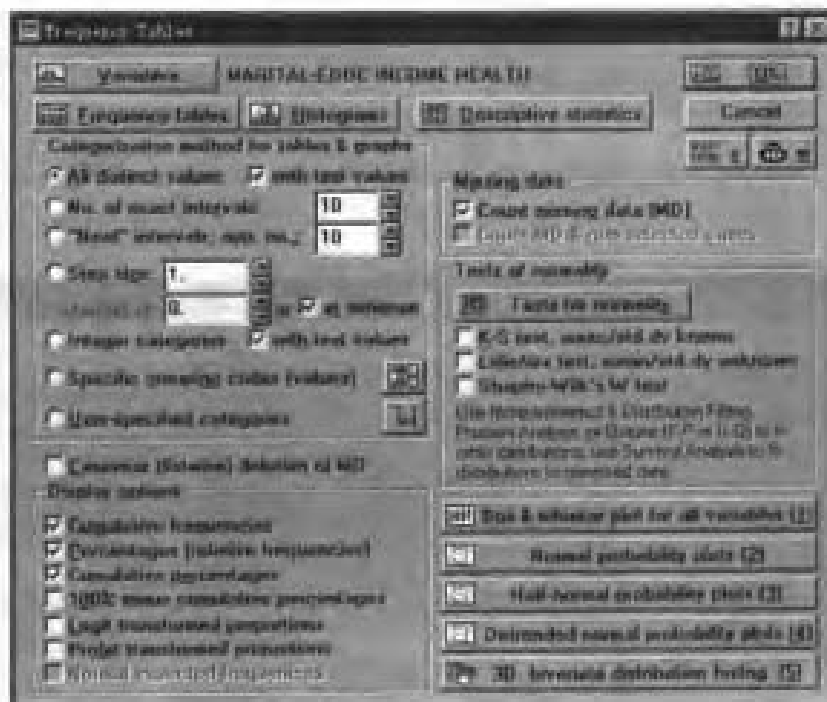


图 45-6-32 Frequency Tables (频数表分析) 主对话框

☆ Percentages [relative frequencies], 百分数 (相关频数), 默认格式。

☆ Cumulative percentages, 累计百分数, 默认格式。

☆ 100% minus cumulative percentages, 100% 负累计百分数。

☆ Logit transformed proportions, Logit 转换比例。

☆ Probit transformed proportions, 概率单位 (Probit, Probability Unit) 转换比例。

☆ Normal expected frequencies, 正态期望频数。

● Missing data, 缺失数据。

☆ Count missing data [MD], 缺失数据 (MD) 个数, 默认格式。

☆ Count MD & non-selected cases, 缺失数据 (MD) 与未选用个案的个数。

● Test of normality, 正态性检验。

☆ K-S test, mean/std. dev. known, 柯尔莫哥罗夫-斯米尔诺夫检验, 已知平均数/标准差。

☆ Lilliefors test, mean/std. dev. unknown, 李尼伏尔检验, 未知平均数/标准差。

☆ Shapiro-Wilk's W test, 夏皮罗-威尔克 W 检验, 用非参数与分布进行拟合, 用批处理分析, 或者用

图形 (P-P 或 Q-Q 图) 拟合其他分布; 用生存分析拟合删失数据的分布。

此外, 还有如下 5 种图形选项:

● Box & whisker plot for all variables [1], 所有变量的箱形与触须线图 (1)。

● Normal probability plots [2], 正态概率图 (2)。

● Half-normal probability plots [3], 半-正态概

率图 (3)。

● Detrended normal probability plots [4], 趋势正态概率图 (4)。

● 3D bivariate distribution histogram [5], 三维二元分布直方图 (5)。

本例的计算结果

(1) 在频数表分析主对话框 (见图 45-6-32) 中, 单击 [Variables] (分析变量), 得到如图 45-6-33 所示对话框, 分析变量可以选择一个或多个, 本例选用数据文件中的多个变量: 如 4-MARITAL, 5-EDUC, 7-INCOME, 32-HEALTH, 然后单击 [OK], 得到如图 45-6-34 所示对话框, 其余选用默认格式。

(2) 再单击 [Frequency tables] (频数分布表), 得到计算结果。

STAT: MARITAL (cond.sts)				
BASIC				
STATS				
		Cumul.		Cumul.
Value	Count	Count	Percent	Percent
1.00000	73	73	24.82993	24.8299
2.00000	127	200	43.19728	68.0272
3.00000	43	243	14.62585	82.6531
4.00000	13	256	4.42177	87.0748
5.00000	38	294	12.92517	100.0000
Missing	0	294	0.00000	100.0000

结果分析:

同样可以得到其他三个分析变量 (5-EDUC, 7-INCOME, 32-HEALTH) 的频数表, 在此不一一列出。

计算结果中给出了各变量值对应的频数 (count), 累计频数 (Cumul. Count), 百分比 (Percent), 累计百分比 (Cumul. Percent)。

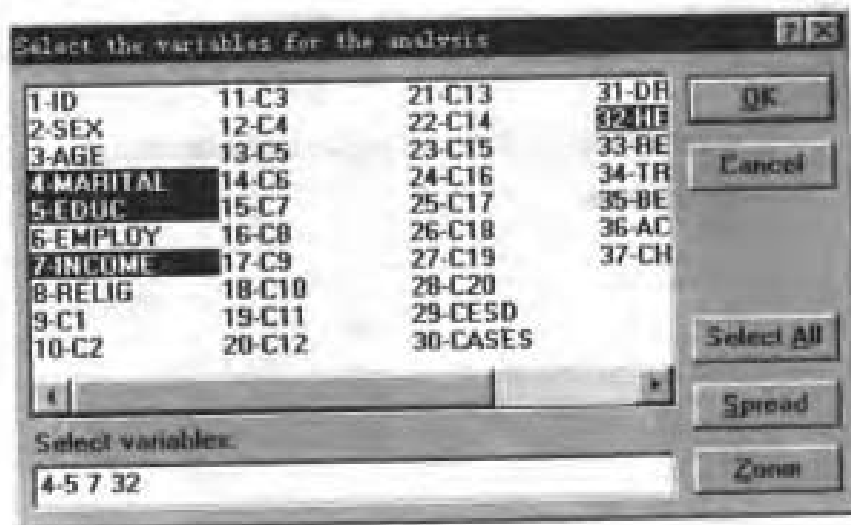


图 45-6-33 选择分析的变量



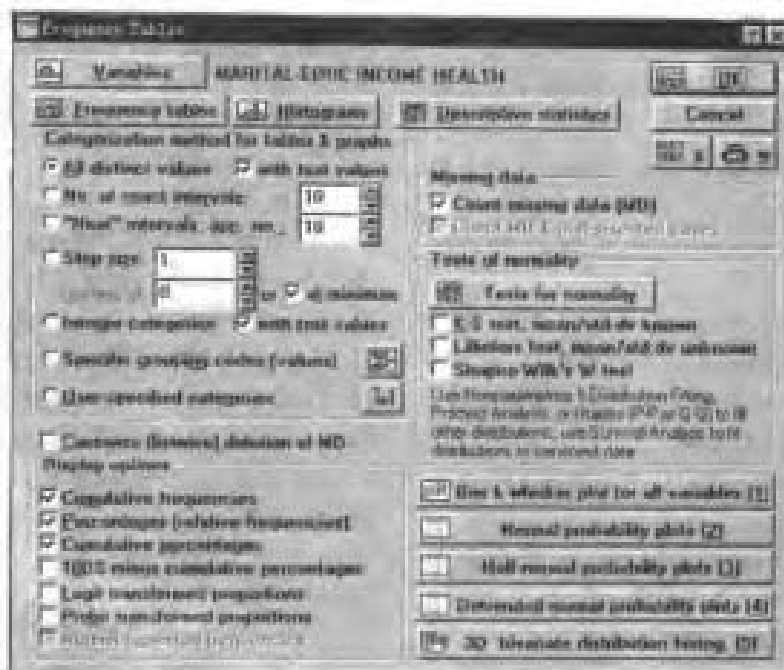


图 45-6-34 已选 Frequency Table (频率分布表) 主对话框

## 5 概率分布计算

概率分布计算器 (Probability Distribution Calculator) 可以对如下 15 种分布函数 (Distribution Function) 与密度函数 (Density Function) 进行计算。它们是贝塔 (Beta) 分布、科西 (Cauchy) 分布、卡方 ( $\chi^2$ , Chi<sup>2</sup>) 分布、指数 (Exponential) 分布、极值 (Extreme value) 分布、F 分布、伽玛 (Gamma) 分布、拉普拉斯 (Laplace) 分布、对数正态 (Log-Normal) 分布、逻辑斯蒂 (Logistic) 分布、帕累托 (Pareto) 分布、瑞利 (Rayleigh) 分布、t (Student, 学生氏) 分布、威布尔 (Weibull) 分布、Z (Normal, 正态) 分布。

当需要计算上述某一个概率分布的值, 或作出其概率分布的图形时, 必须先进入概率分布计算器 (Probability Distribution Calculator) 模型, 然后, 选

用适当的选项, 键入相应的参数。

进入 STATISTICA/Win, 单击 [Basic Statistics/Tables] ⇒ [Switch To] ⇒ [Probability Distribution Calculator] (概率分布计算器) ⇒ [OK], 得到图 45-6-35。

概率分布计算器 (Probability Distribution Calculator) 主对话框的左侧, 显示概率分布的 15 种类型, 默认是贝塔 (Beta) 分布。左侧下方是 Fixed Scaling, 固定尺度, 默认选用此项。备选项有:

- Inverse, 逆贝塔 (Beta) 分布。
- Two-tailed, 双侧。
- [1-Cumulative p], 1-累积 p 值。
- Print, 打印。
- Create Graph, 建立图形。
- Beta, (贝塔值)。

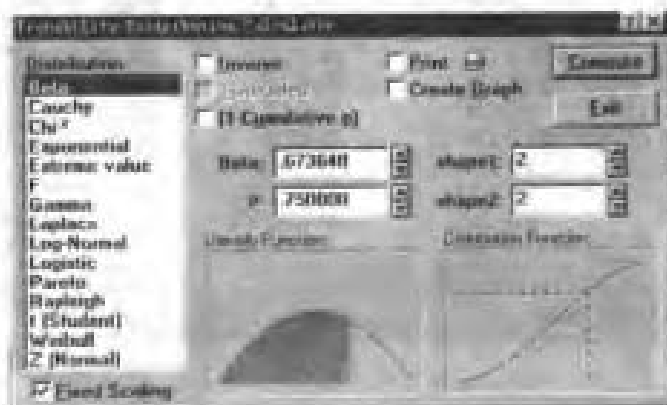


图 45-6-35 Probability Distribution Calculator (概率分布计算器) 主对话框

- $p$ , ( $p$  值),
- shape1, 形状 1,
- shape2, 形状 2,

右下方是贝塔 (Beta) 分布的概率密度函数 (Density Function) 与贝塔 (Beta) 分布函数 (Distribution Function) 的图形。下面仅以贝塔 ( $\beta$ , Beta) 分布、Z (Normal, 正态) 分布、t (Student, 学生氏) 分布为例, 予以介绍, 其他分布的计算方法类似。

### 5.1 贝塔 (Beta) 分布

Beta (quant; shape1, shape2), 其中, quant 是数

量, shape1 是形状 1, shape2 是形状 2。

试计算贝塔 (Beta) 分布函数 Beta (0.5, 5, 3) 的值并作图形。

在图 45-6-36 所示对话框中, 选用 Fixed Scaling (固定尺度), Print (打印) 与 Create Graph (建立图形), 在 Beta 中, 键入 0.5, 在 shape1 中, 键入 5, 在 shape2 中, 键入 3, 如图所示, 再单击 [Compute] (计算), 得到结果:

$$B(0.5, 5, 3) = 0.226563 (P)$$

和贝塔概率密度函数 (Probability Density Function) 图形 (见图 45-6-37)。

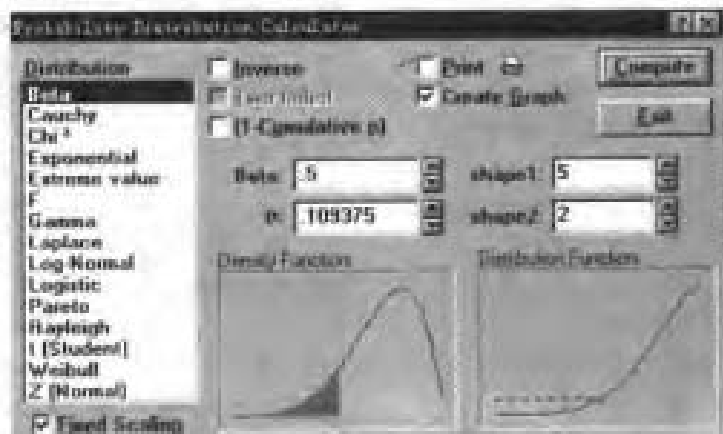


图 45-6-36 Probability Distribution Calculator (贝塔概率分布计算器) 主对话框

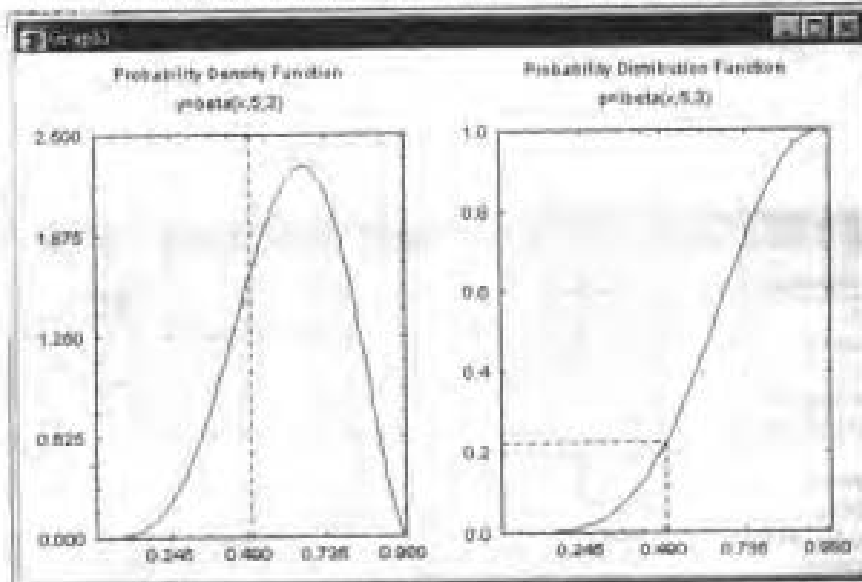


图 45-6-37 贝塔概率密度函数 (Probability Density Function) 与分布函数 (Probability Distribution Function) 并作图。

### 5.2 Z (Normal) 分布

Z (quant, mean, st.dev.), 其中 quant 是数量, mean 是平均数, st.dev. 是标准差。

例 试计算 Z (Normal) 分布 Z (0, 0, 1) 的值

在图 45-6-35 中, 分布函数选用 Z (Normal), 在 Z 中, 键入 0; 在 mean 中, 键入 0; 在 st.dev. 中, 键入 1, 再单击 [Compute], 得到如图 45-6-38 所示的 Z (Normal) 概率分布计算界面, 得到  $Z(0, 0, 1) =$

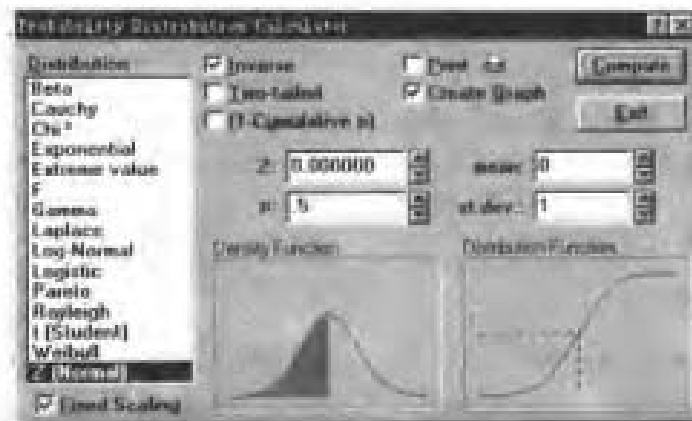


图 45-6-38 Z (Normal) 概率分布计算器

0.000000, Z (Normal) 的概率密度函数与概率分布函数图形如图 45-6-39 所示。

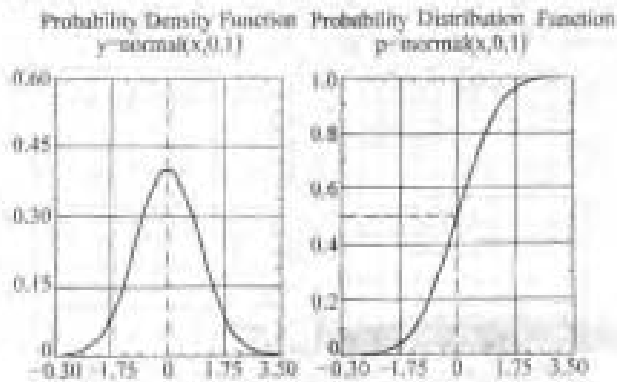


图 45-6-39 Z (Normal) 概率密度函数与概率分布函数

### 5.3 t (Student) 分布

$t(t, df)$ , 其中,  $t$  是数量,  $df$  是自由度。

例 试计算学生氏 ( $t$ ) 分布函数  $t(2.539373, 22)$  的值并作图。

在图 45-6-35 中, 分布函数选用  $t$  (Student), 在  $t$  中, 键入 2.539373; 在  $df$  中, 键入 22, 选用: 双侧概率 (Two-tailed), 建立图形 (Create Graph) 与 “1-累积 P 值”, 再单击 [Compute] 得到如图 45-6-40 所示的概率分布计算器, 并得到  $t(2.539373, 22) = 0.018684$ ,  $t$  (student) 的概率密度函数与概率分布函数图形, 如图 45-6-41 所示。

用类似的方法还可以对其他概率分布函数与概率密度函数进行计算并作图。

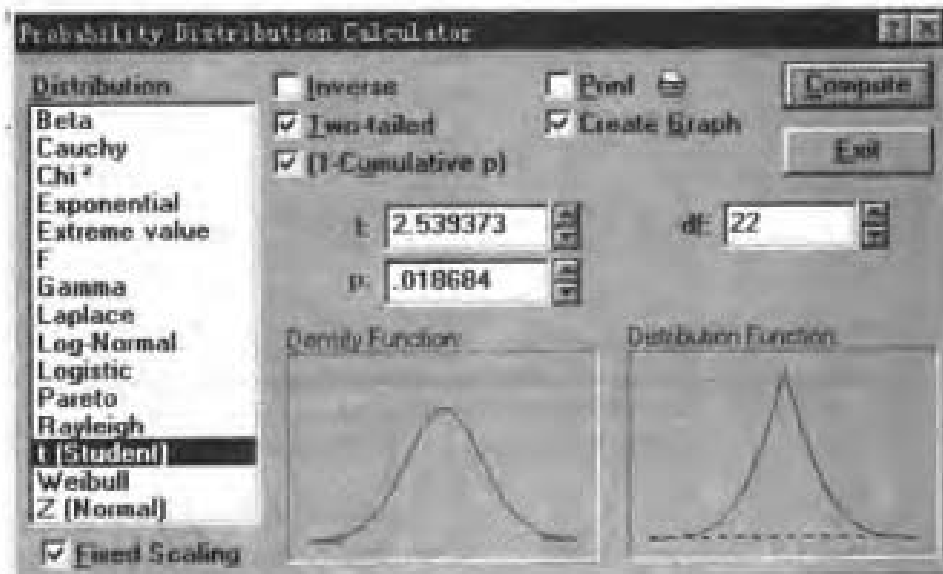


图 45-6-40 t 概率分布计算器

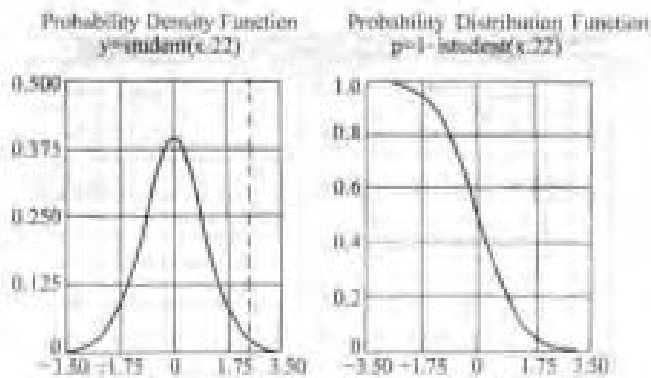


图 45-6-41 t (Student) 概率密度函数  
与概率分布函数

## 6 非参数统计检验/分布拟合

非参数统计检验/分布拟合 (Nonparametrics/Distrib) 包括: ①非参数统计检验 (Nonparametric Stats), 它有  $2 \times 2$  表 ( $2 \times 2$  Table),  $\chi^2/P/\Phi$  值, 麦克内马检验 (McNemar), 费歇尔精确检验 (Fisher Exact), 观察值与期望值的  $\chi^2$  检验 (Observed Versus Expected  $\chi^2$ ), 非参数相关分析 (Correlations (Spearman, Kendall, Gamma)), 沃尔德-沃尔福owitz游程检验 (Wald-Wolfowitz Runs Test), 曼-惠特尼 U 检验 (Mann-Whitney U tests), 柯尔莫哥罗夫-斯米尔诺夫 2 样本检验 (Kolmogorov-Smirnov Two-sample Test), 克鲁斯凯-沃里斯方差分析 (中位数检验) (Kruskal-Wallis ANOVA, mediantest), 符号检验 (Signtest), 威尔科克森配对检验 (Wilcoxon-MatchedPairstest), 弗里德曼方差分析与肯德尔一致性 (Friedman ANOVA & Kendall's Concordance), 科克伦 Q 检验 (Cochran Qtest) 等。②分布拟合 (Distribution Fitting), 它能对连续型分布 (Continuous Distributions) 如正态分布 (Normal), 均匀分布 (Rectangular), 指数分布 (Exponential),  $\Gamma$  分布 (Gamma), 对数正态分布 (Log-normal),  $\chi^2$  分布 (Chi-square) 及其他 (Others) 连续型分布进行拟合, 又能对离散型分布 (Discrete Distributions) 如二项分布 (Binomial), 泊松分布 (Poisson), 几何分布 (Geometric) 与伯努利分布 (Bernoulli) 进行拟合。

### 6.1 常用描述性统计分析

常用描述性统计分析 (Ordinal descriptive statistics (median, mode, ...)) 可以计算一个或多个统计指标, 它包括均值 (mean), 有效例数 (valid N), 中位数 (median) 与指定的百分位数等。

例 试对某数据文件: ccsd.sta 中的 income 分析变量做常用描述性统计分析, 并指定百分位数是: 5th, 95th。

- 1) 调入数据文件: ccsd.sta。
- 2) 进入 STATISTICA/Win, 单击 [Nonparametrics/Distrib] (非参数统计检验/分布拟合)  $\Rightarrow$  [Switch To]  $\Rightarrow$  [Ordinal] descriptive statistics (median, mode, ...) (常用描述性统计分析)  $\Rightarrow$  [OK], 得到如图 45-6-42 所示对话框。

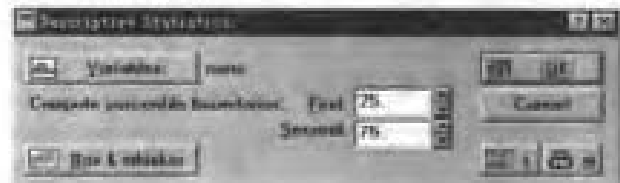


图 45-6-42 Descriptive Statistics (描述性统计分析) 对话框

- 3) 在图 45-6-42 所示对话框中, Variables 是分析变量; Compute percentile boundaries, 计算百分位数的边界; First, 指定第 1 个百分位数 (默认是 25th); Second, 指定第 2 个百分位数 (默认是 75th)。
- 4) 单击 [OK], 得到如图 45-6-43 所示对话框。



图 45-6-43 选择变量

- 5) 在图 45-6-43 所示对话框中, 选用变量 7-INCOME, 其数据为 4, 15, 28, 9, 35, 11, 31, 9, 23, 25, 25, 24, 28, 13, 15, 68, 19, 15, 9, 6, 35, 7, 19, 6, 19, 13, 5, 19, 20, 19, 45, 35, 23, 3, 15, 19, 15, 23, 23, 11, 23, 35, 28, 23, 9, 35, 35, 55, 11, 5, 28, 6, 7, 23, 6, 20, 23, 15, 28, 6, 35, 23, 8, 9, 28, 11, 15, 19, 4, 45, 13, 2, 7, 12, 12, 65, 7, 11, 7, 5, 4, 7, 13, 23, 35, 11, 13, 28, 19, 2, 8, 15, 19, 63, 7, 8, 4, 11, 45, 19, 8, 65, 35, 36, 13, 7, 18, 23, 27, 9, 15, 23, 2, 23, 19, 15, 13, 8, 9, 65, 9, 19, 2, 4, 9, 7, 9, 15, 15, 19, 11, 7, 8, 13, 35, 15, 45, 45, 2, 9, 13, 45, 20, 28, 35, 11, 9, 19, 28, 7, 13, 9, 8, 5, 11, 19, 16, 35, 32, 35, 55, 42, 5, 36, 65, 45, 55, 19, 8, 19, 9, 5, 19, 5, 4, 6, 35, 23, 23, 15, 19, 9, 19, 6, 7, 23, 7, 13, 5, 15, 9, 2, 45, 4, 24, 45, 9, 26, 23, 19, 7, 19, 19, 6, 6, 28, 28, 23, 23, 23, 28,

6, 8, 5, 15, 11, 19, 35, 11, 45, 65, 65, 55, 28, 13, 31, 7, 15, 8, 55, 45, 35, 65, 11, 55, 7, 13, 19, 45, 45, 37, 23, 19, 7, 15, 15, 9, 28, 8, 7, 9, 9, 55, 15, 2, 11, 28, 15, 35, 15, 15, 13, 9, 23, 9, 15, 15, 23, 65, 35, 13, 15, 11, 11, 8, 6, 23, 6, 35, 45, 35, 28, 4, 13, 35, 35, 28, 28, 23, 35, 55, 28, 9 单击 [OK], 得到图 45-6-44。First, 指定第 1 个百分位数 (默认是 25th) 选用 5, Second, 指定第 2 个百分位数 (默认是 75th)。

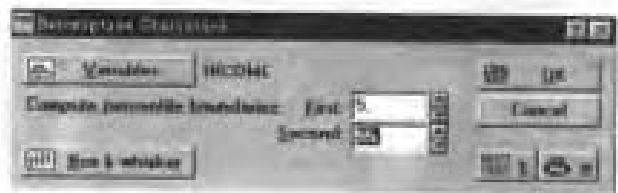


图 45-6-44 已选 Descriptive Statistics (描述性统计分析) 对话框  
6) 单击 [OK], 得到结果。

STAT-	Descriptive Statistics (readl.sta)							
NONPAR								
STATS								
variable	mean	valid N	median	mode	freq. of mode	minimum		
变量	均值	有效例数	中位数	众数	众数的频数	最小值		
INCOME	20.07483	294	15.00000	multiple		2.000000		
STAT-	Descriptive Statistics (readl.sta)							
NONPAR								
STATS								
variable	maximum	6.0000th percentil	75.0000th percentil	geometric mean	harmonic mean	std. dev.		
变量	最大值	第 5 百分位数	第 75 百分位数	几何均数	调和均数	标准差		
INCOME	65.00000	4.000000	55.00000	15.56678	11.31228	18.89012		
STAT-	Descriptive Statistics (readl.sta)							
NONPAR								
STATS								
variable	variance	average derivati.	range	quartile range	skewness	kurtosis		
变量	方差	平均差	全距	分位数距	偏度	峰度		
INCOME	233.7879	12.01773	63.00000	19.00000	1.229047	.558304		
STAT-	Descriptive Statistics (readl.sta)							
NONPAR								
STATS								
variable	sum							
变量	总和							
INCOME	6049.100							

### 模型特色

本模型的最大特色, 可以由用户任意选定百分位数, 而同时又可以计算其他常用描述性统计量, 分析的变量可以是多个。

## 6.2 观察与期望频数的对照检验

观察与期望频数的对照检验 (Observed versus expected frequency) 属于  $\chi^2$  检验的单向分类 (Oneway classification)。在  $\chi^2$  检验中的基本变量是离散 (分类) 变量, 而在单向分类中, 仅有一个离散变量。

例 有人猜测卡尔·马克思在其写作中, 常常等概率地使用 a, e, i, o, s 字母。为了检验这个猜测, 从他的作品中, 随机地抽取了一些字母, 上述 5 个字母的出现频数见表 45-6-6。共抽取了 500 个字母, 如果符合均匀分布, 则各字母的理论频数应为  $500/5=100$ , 设数据文件建于 d:\stadata\obsexp.Sta, 试检验

这个猜测是否正确。

表 45-6-6 卡尔·马克思在其写作中字母 a, e, i, o, s 的频数

字母	a	e	i	o	s	合计
观察频数	90	115	105	84	106	500
理论频数	100	100	100	100	100	100

1) 进入 STATISTICA/Win, 单击 [Nonparametric/Distrib.], (非参数统计检验/分布拟合)  $\Rightarrow$  [Switch To]  $\Rightarrow$  [Observed versus expected N (观察与期望频数的对照检验)]  $\Rightarrow$  [OpenData (d:\stadata\obsexp.sta)], 得到如图 45-6-45 所示数据文件, 再单击 [OK], 得到如图 45-6-46 所示对话框。

2) 在图 45-6-46 所示对话框中单击 [Variables], 得到如图 45-6-47 所示对话框。

3) 在图 45-6-47 所示对话框中, 观察频数

Variable	1 OBS	2 EXP
1	90	100
2	115	100
3	105	100
4	84	100
5	106	100

图 45-6-46 数据文件 obsexp.sta



图 45-6-48 已选 Observed vs. Expected Frequency (观察与期望频数) 对话框

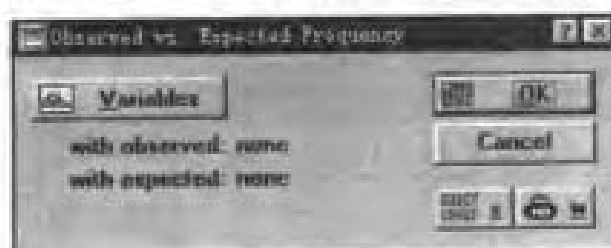


图 45-6-49 Observed vs. Expected Frequency (观察与期望频数) 对话框

(Observed frequencies) 是 1-OBS, 期望频数 (Expected frequencies) 是 2-EXP, 单击 [OK], 得到如图 45-6-48 所示对话框, 再在该对话框中单击 [OK] 得到结果。

STAT: Observed vs. Expected Frequencies(obsexp.sta)  
NONPAR Chi-Square=6.42000 df=4 p<.169950

STATS

Case	observed - expected		(O-E) / √E	
	OBS	EXP	O-E	√E
C <sub>1</sub>	90.0000	100.0000	-10.0000	1.00000
C <sub>2</sub>	115.0000	100.0000	15.0000	2.25000
C <sub>3</sub>	105.0000	100.0000	5.0000	2.25000
C <sub>4</sub>	84.0000	100.0000	-16.0000	2.50000
C <sub>5</sub>	106.0000	100.0000	6.0000	2.36000
Sum	500.0000	500.0000	0.0000	6.42000

### 结果分析

由于  $\chi^2$  (Chi-Square) = 6.420,  $p < 0.169950$ , 尚不能得到卡尔·马克思写作中的 5 个字母不等概率出现的结论。换一句话说, 猜测可能正确。

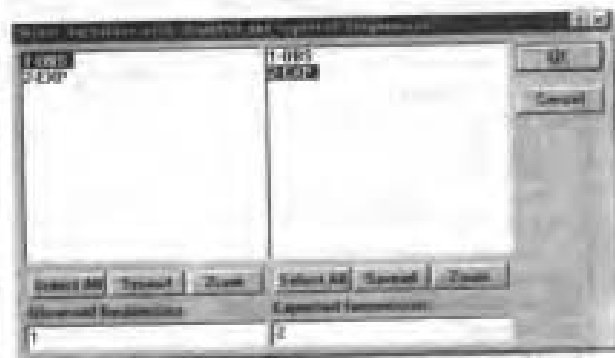


图 45-6-47 选择变量

## 6.3 柯尔莫哥罗夫-斯米尔诺夫两样本检验

STATISTICA/Win 中的柯尔莫哥罗夫-斯米尔诺夫两样本检验 (Kolmogorov-Smirnov two-sample test) 可以对两独立样本进行检验, 用以判断两独立样本的分布是否存在显著性差异。

例 观察材料 1 的 5 个试样与材料 2 的 13 个试样的硬度值, 结果如表 45-6-7, 试判断材料 1 的试样与材料 2 的试样的硬度值的差异是否具有显著性意义。

表 45-6-7 材料 1 的试样与材料 2 的试样的硬度值

材料 1	25	34	44	48	49								
材料 2	15	15	16	17	19	21	23	25	27	28	28	30	35

1) 建立数据文件: k-s2.sta (见图 45-6-49)。

图 45-6-49 所示界面中, 第 1 个变量是 1-G (分组变量), 其中第 1 组变量是材料 1, 第 2 组变量是材料 2。第 2 个变量是 1-T (试样的硬度值)。

Variable	1	2
1	1	25.0
2	1	34.0
3	1	44.0
4	1	48.0
5	1	49.0
6	2	15.0
7	2	15.0
8	2	16.0
9	2	17.0
10	2	19.0
11	2	21.0
12	2	23.0
13	2	25.0
14	2	27.0
15	2	28.0
16	2	28.0
17	2	30.0
18	2	35.0

图 45-6-49 数据文件: k-s2.sta

2) 单击 [Nonparametrics/Distrib.] (非参数统计检验分布拟合) ⇒ [Switch To] ⇒ [Kolmogorov-Smirnov two-sample test] ⇒ [OK], 单击 [Variables] (分析变量), 自变量 (分组变量) (Independent (Grouping) variable) 选用 1-G, 因变量清单 (Dependent variables list) 选用 1-T, 再单击 [OK], 得到图 45-6-50 所示对话框, 单击 [OK], 得到结果。

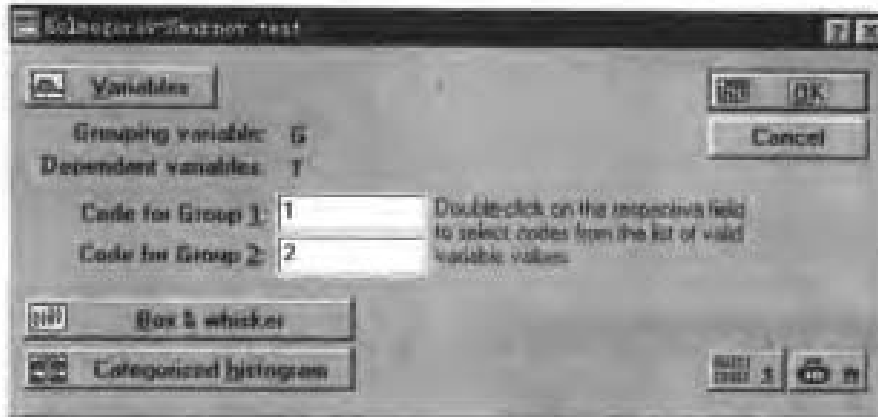


图 45-6-50 Kolmogorov-Smirnov test (柯尔莫哥罗夫-斯米尔诺夫大检验) 对话框

STAT.	Kolmogorov-Smirnov Test (k-s2.sta)						
NONPAR	By variable G						
STATS	Group1: 1Group2: 2						
variable	Max Neg Differe	Max Pos Differe	p-level	Mean Group1	Mean Group2	Std. Dev. Group1	
变量	最大负差异	最大正差异	p-水准	1组均值	2组均值	1组标准差	
T	0.00	.728571	p<.05	39.00000	22.85714	9.273619	
STAT.	Kolmogorov-Smirnov Test (k-s2.sta)						
NONPAR	By variable G						
STATS	Group1: 1Group2: 2						
variable	Std. Dev. Group2	Valid N Group1	Valid N Group2				
	2组标准差	1组有效例数	2组有效例数				
T	6.109966	5	14				

**结果分析**

由于  $p < 0.05$ ，材料 1 的试样与材料 2 的试样的硬度值分布是不同的。

**6.4 分布拟合**

分布拟合 (Distribution Fitting) 包含在非参数统计检验分布拟合 (Nonparametric/Distrib.) 中，前面介绍了非参数统计检验 (Nonparametric Stats) 的常用方法，下面分别介绍拟合连续型分布 (Continuous Distribution Fitting) 和离散型分布 (Discrete Distribution)。

**6.4.1 拟合连续型分布**

拟合连续型分布能进行正态分布 (Normal)，均匀分布 (Rectangular)，指数分布 (Exponential)， $\Gamma$  分布 (Gamma)，对数正态分布 (Log-normal)， $\chi^2$  分布 (Chi-square) 与其他 (Others) 连续型分布等分布的拟合。

(1) 正态分布 (Normal) (计量资料)

例 某实验对 100 个失效试样测定了断面面积，如表 45-6-8，试拟合—正态分布。

1) 进入 STATISTICA/Win，建立数据文件，request.stad<sub>1</sub> (d)stadata。

2) 单击 [Nonparametrics/Distrib.] (非参数统计检验分布拟合) ⇒ [SwitchTo]，得到如图 45-6-51 所

示对话框。

表 45-6-8 失效试样断面面积数据

74.3	78.8	88.8	78.0	70.4	80.5	80.5	69.7	71.2	73.5
79.6	76.6	75.0	78.8	72.0	72.0	72.0	74.3	71.2	72.0
75.0	74.5	78.8	78.3	72.8	65.0	74.3	71.2	69.7	68.0
77.5	75.0	72.0	84.3	72.8	80.3	69.7	74.3	73.5	73.5
75.8	76.8	88.8	76.5	70.4	71.2	81.2	75.0	70.4	68.0
74.0	72.0	78.5	74.3	76.3	77.8	87.3	72.0	75.0	74.3
73.5	78.5	73.5	74.7	65.0	76.8	81.6	75.4	72.7	72.7
69.2	76.5	72.7	70.4	72.2	68.8	67.3	67.3	67.3	72.7
75.8	73.5	75.0	72.7	73.5	13.5	72.7	81.6	70.3	74.8
72.5	79.5	70.4	76.5	72.7	77.2	84.3	75.0	76.5	70.4

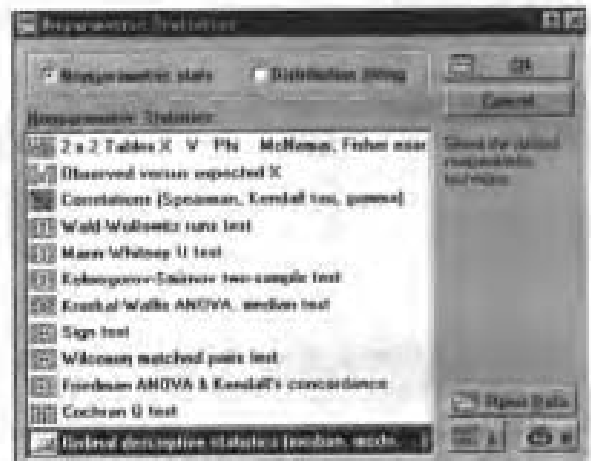


图 45-6-51 Nonparametric Statistics (非参数统计) 对话框

3) 在图 45-6-51 所示对话框中, 选用 [Distribution fitting] (分布拟合), 得到图 45-6-52 所示对话框, 可选项如下:



图 45-6-52 Distribution Fitting (分布拟合) 主对话框

● Continuous Distributions: 连续型分布。它包括:

- ☆ Normal, 正态分布。
- ☆ Rectangular, 均匀分布。
- ☆ Exponential, 指数分布。
- ☆ Gamma, 伽玛 $\Gamma$ 分布。
- ☆ Log-normal, 对数正态分布。
- ☆ Chi-square,  $\chi^2$  分布。
- ☆ Others, 其他连续型分布。

● Discrete Distributions: 离散型分布。它包括:

- ☆ Binomial, 二项分布。
- ☆ Poisson, 泊松分布。
- ☆ Geometric, 几何分布。
- ☆ Bernoulli, 伯努利分布。

4) 本例选用正态分布, 单击 [Normal]  $\Rightarrow$  [OK], 得到如图 45-6-53 所示对话框。

图 45-6-53 所示界面的可选项如下。

- Distribution: (待选) 分布, 默认是 Normal
- ☆ Variable: none, 拟合分布的变量。
- ☆ Number of categories: 分类数。
- ☆ Lower limit: 下限。
- ☆ Upper limit: 上限。
- ☆ Mean: 均值。
- ☆ Variance: 方差。
- Kolmogorov Smirnov test: 柯尔莫哥罗夫-斯米尔诺夫检验。
- ☆ No: 不做检验。
- ☆ Yes (categorized): 做检验 (分类), 默认选用此项。

☆ Yes (continuous): 做检验 (连续型)。

● Graph: 图形。

☆ Graph: 直方图。

☆ Plot distribution: 图形的分布。

Frequency distribution: 频数分布。

Cumulative distribution: 累计分布。

☆ Plot raw frequencies or %: 原始频数或百分数 (%) 图形。

Raw frequencies: 原始频数。

Relative frequencies (%): 相对频数 (%)。

5) 在图 45-6-53 所示对话框中, 单击 [Variable], 得到如图 45-6-54 所示界面。

6) 在图 45-6-54 所示界面中, 选用变量 I-SERUM, 然后单击 [OK], 得到如图 45-6-55 所示的对话框。其中有如下选项数值。

● 分类数 (Number of categories) = 13,

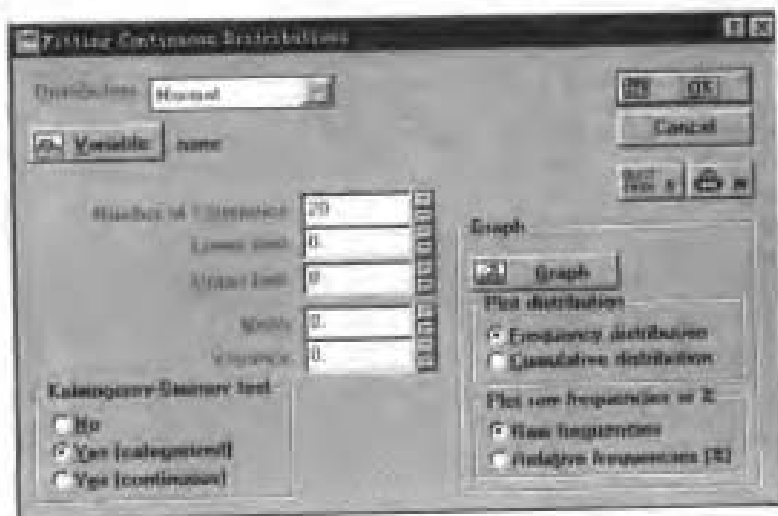


图 45-6-53 Fitting Continuous Distributions (拟合连续分布) 主对话框





图 45.6-54 选择变量

- 下限 (Lower limit) = 62.0,
- 上限 (Upper limit) = 88.0,
- 均值 (Mean) = 73.696014,
- 方差 (Variance) = 15.416745.

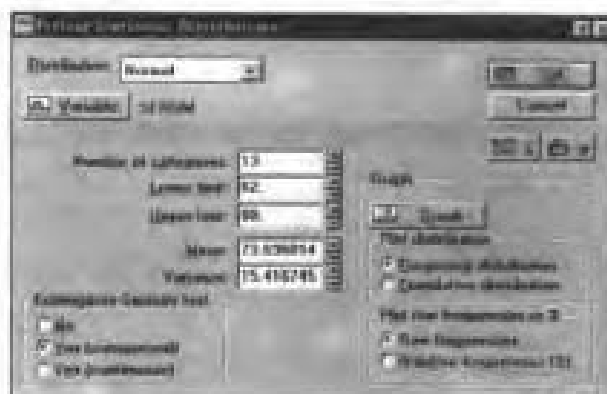


图 45.6-55 已选 Fitting Continuous Distributions 拟合正态分布主对话框

此外,在图 45.6-55 中,单击 [OK],可得到结果。

STAT-	Variable SERUM, distribution: Normal (frequentl-sta)						
NONPAR	K-S d = .0700923, p = n, s, Lilliefors p = n, s						
STATS	Chi-Square: 5.457545, df = 5, p = .3626468 (df adjusted)						
	z 值	p 值					
Upper Boundary	observed	cumulative	percent	cumul. %	expected	cumulative	
上边界	freq-cy	observed	observed	observed	freq-cy	expected	expected
	观察频数	累计观察数	观察百分比	累计观察百分数	期望频数	累计期望数	
<= 64	0	0	0.00000	0.0000	-.67664	.6766	
66	2	2	2.00000	2.0000	1.82279	2.4994	
68	7	10	7.00000	10.0000	4.84387	7.3433	
70	9	16	6.00000	16.0000	9.98374	17.3270	
72	18	34	18.00000	34.0000	15.96185	33.2889	
74	19	53	19.00000	53.0000	19.79686	53.0858	
76	23	76	23.00000	76.0000	19.04724	72.1328	
78	11	87	11.00000	87.0000	14.21680	86.3496	
80	8	93	8.00000	93.0000	8.23157	94.5812	
82	6	99	6.00000	99.0000	3.69693	98.2781	
84	0	99	0.00000	99.0000	1.28774	99.5658	
86	1	100	1.00000	100.0000	-.34785	99.9137	
Infinity	0	100	0.00000	100.0000	.08633	100.0000	
无限大							

STAT-	Variable SERUM, distribution: Normal (frequentl-sta)			
NONPAR	K-S d = .0700923, p = n, s, Lilliefors p = n, s			
STATS	Chi-Square: 5.457545, df = 5, p = .3626468 (df adjusted)			
Upper Boundary	percent	cumul. %	observed-	
上边界	expected	expected	observed-	
	期望百分数	累计期望百分数	观察数-期望数	
<= 64	.67664	.6766	-.67664	
66	1.82279	2.4994	1.17721	
68	4.84387	7.3433	2.15413	
70	9.98374	17.3270	-5.98374	
72	15.96185	33.2889	3.03816	
74	19.79686	53.0858	-.79686	
76	19.04724	72.1328	3.95279	
78	14.21680	86.3496	-3.21680	
80	8.23157	94.5812	-2.23157	
82	3.69693	98.2781	2.30907	
84	1.28774	99.5658	-1.28774	
86	-.34785	99.9137	+.6219	
Infinity	.08633	100.0000	-.08633	

## 图形结果

在图 45.6-55 所示对话框中,单击 [Graph],得到如图 45.6-56 所示的图形。横轴是断面面积,纵轴是观察数 (No. of obs)。

## 结果分析

1) 由于  $\chi^2$  值 (Chi-Square) = 5.457543,  $P (P = 0.3626468) > 0.05$ ,按  $\alpha = 0.05$  水准,可认为观察频数与期望频数符合较好,服从正态分布。

2) 又由于正态分布曲线是以均数为中心呈对称的钟形分布。本例由图 45.6-55 所示图形结果也可看出,它基本上是以均数 (7.696014) 为中心,中心凸出,两侧对称的钟形图形。

正态分布 (Normal) 又称高斯 (Gauss) 分布,它是一种很重要的连续型分布,有些分布,在一定条件下也可近似用正态分布 (Normal) 描述。因而它在各种自然科学或社会科学中,有很广泛的理论与实际应用价值。

Variable SERUM; distribution: Normal  
Kolmogorov-Smirnov d = 0.096720, p = n.s.  
Chi-Square = 5.457543, df = 5, p = 0.3626468 (df adjusted)

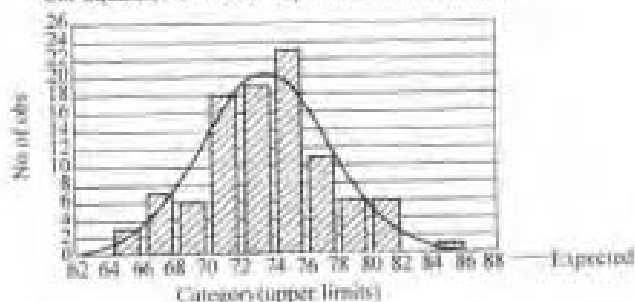


图 45.6-56 正态分布图 (期望正态分布曲线与直方图)

## (2) 正态分布 (Normal) (频数表资料)

当观察数较多时,即大样本资料,有时先整理成频数表,然后再做拟合或其他检验。

例 110 个试样直径的频数表资料如表 45.6-9,试拟合一正态分布曲线。其中, X 直径组中值, F 是频数。

本例原始的数据是频数表资料,应先对频数 F 进行加权,然后才能对 X 拟合正态分布曲线。

1) 进入 STATISTICA/Win, 建立数据文件: (frequenz.sta (d:\stadata)) (见图 45.6-57)。

	X	F
1	10.75	1
2	11.25	3
3	11.75	1
4	12.25	5
5	12.75	7
6	13.25	13
7	13.75	16
8	14.25	19
9	14.75	18
10	15.25	14
11	15.75	8
12	16.25	4
13	16.75	2
14	17.25	1

图 45.6-57 数据文件: frequenz.sta

2) 单击 [Options]  $\Rightarrow$  [Weight] (加权), 得到图 45.6-58 所示的对话框。

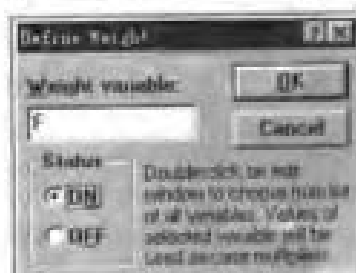


图 45.6-58 Define Weight (定义权重变量)

权重变量 (Weight variable) 是 F, 变量状态 (Status) 是 ON, 单击 [OK] (完成了对 F 的加权)。

3) 单击 [Nonparametric/Distrib.] (非参数统计检验/分布拟合)  $\Rightarrow$  [Switch To] (见图 45.6-51)  $\Rightarrow$  [Distribution fitting] (分布拟合) (图 45.6-52)  $\Rightarrow$  [Normal] (正态分布)  $\Rightarrow$  [OK] (图 45.6-53)  $\Rightarrow$  [Variable] (待选变量)  $\Rightarrow$  [Select variable] (选择变量) 是 1-X  $\Rightarrow$  [OK], 得到如图 45.6-59 所示界面。

4) 单击 [OK], 得到计算结果。

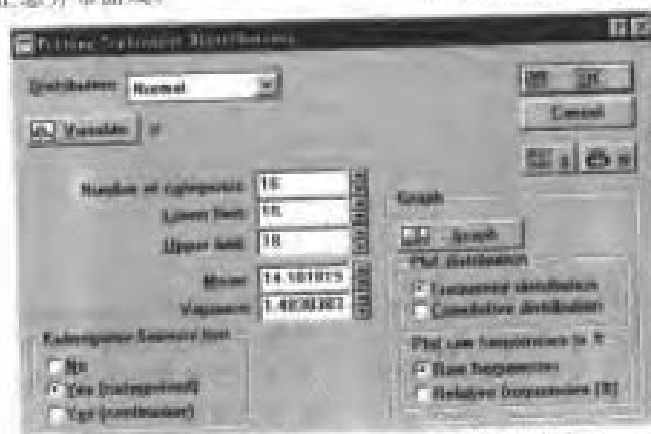


图 45.6-59 已选 Fitting Continuous Distributions (拟合正态分布 Normal) 主对话框

表 45.6-9 试样直径的频数表资料

X	10.75	11.25	11.75	12.25	12.75	13.25	13.75	14.25	14.75	15.245	15.75	16.25	16.75	17.25
F	1	1	3	5	7	13	16	19	18	14	6	4	2	1

data file: FREQUEN2.STA [14cases with 2 variables]

Var1: X-组中值 (-9999)

Distribution: Normal

mean: 14.18182

variance: 1.483838

STAT. Variable X; distribution: Normal (frequen2.sta)

NONPAR Kolmogorov-Smirnov d=.0224921, p=n.s.

STATS Chi Square: 1.627110, df=6, p=.9506042 (df adjusted)

Upper Boundary	observed freq-cy	cumulatv observed	percent observed	cumul. % observed	expected freq-cy	cumulatv expected
上边界	观察频数	累计观察数	观察百分数	累计观察百分数	期望频数	累计期望数
<=10.	0	0	0.00000	0.0000	.13787	.1379
11.0	1	1	.90909	.9091	.35713	.4950
11.5	1	2	.90909	1.8182	1.02819	1.5232
12.0	3	5	2.72727	4.5455	2.50685	4.0300
12.5	5	10	4.54545	9.0909	5.17609	9.2061
13.0	7	17	6.36364	15.4545	9.05116	18.2573
13.5	13	30	11.81818	27.2727	13.40434	31.6616
14.0	16	46	14.54545	41.8182	16.81250	48.4741
14.5	19	65	17.27273	59.0909	17.85946	66.3336
15.0	18	83	16.36364	75.4545	16.06776	82.4014
15.5	14	97	12.72727	88.1818	12.24308	94.6444
16.0	6	103	5.45455	93.6364	7.90078	102.5452
16.5	4	107	3.63636	97.2727	4.31804	106.8633
17.0	2	109	1.81818	99.0909	1.99861	108.8619
17.5	1	110	.90909	100.0000	.78340	109.6453
Infinity	0	110	0.00000	100.0000	.35473	110.0000

无限大

STAT. Variable X; distribution: Normal (frequen2.sta)

NONPAR Kolmogorov-Smirnov d=.0224921, p=n.s.

Upper Boundary	percent expected	cumul. % expected	observed expected
上边界	期望百分数	累计期望百分数	观察数-期望数
<=10.	.12534	.1253	-.13787
11.0	.32466	.4500	-.64287
11.5	.934721	.3847	-.02819
12.0	2.27896	3.6637	-.49315
12.5	4.70554	8.3692	-.17609
13.0	8.22833	16.5975	-2.05116
13.5	12.18576	28.7833	-.40434
14.0	15.28409	44.0674	-.81250
14.5	16.23588	60.3033	1.14054
15.0	14.60705	74.9103	1.93224
15.5	11.13007	86.0404	1.75692
16.0	7.18253	93.2229	-1.90078
16.5	3.92549	97.1484	-3.1804
17.0	1.81692	98.9653	.00139
17.5	.71218	99.6775	.21660

图形结果

在图 45.6-59 所示界面中,单击 [Graph],得到如图 45.6-60 所示的图形结果。

在图 45.6-60 中横轴是直径 (X),纵轴是观察数

(No of obs)。本例 P (p=0.9506042) > 0.95,直方图与正态分布理论曲线 (Expected) 拟合也很吻合,表明拟合优度较好。

类似的方法,可以做其他连续型分布的拟合。

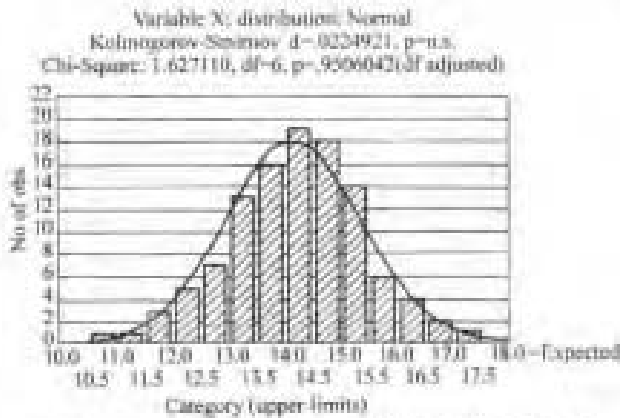


图 45.6.60 正态分布图 (期望正态分布曲线与直方图)

### 6.4.2 拟合离散型分布

分布拟合能对离散型分布 (Discrete Distributions) 进行拟合。例如, 二项分布 (Binomial)、泊松分布 (Poisson)、几何分布 (Geometric) 与伯努利分布 (Bernoulli)。

#### 1. 二项分布 (Binomial)

例 已知从某加工的试件中, 随机抽取 50 件, 每 10 件为一组, 其中合格的试件数如表 45.6.10。问此资料是否服从二项分布。

表 45.6.10 每 10 件试件中的合格数

每 10 件试件中的合格数, X	0	1	2	3	4	5	6	7	8	9	10
例数, F		2	2	9	12	12	9	3	1		

1) 建立数据文件: binomial.vta (见图 45.6.61)。

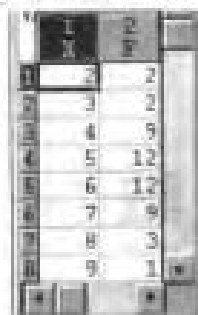


图 45.6.61 数据文件

2) 单击 [Options] → [Weight] (对 F 加权)。

3) 单击 [Nonparametrics/Distrib.] (非参数统计检验分布拟合) → [Switch To] → [Distribution fitting] (分布拟合) → [Binomial] (二项分布), 得到如图 45.6.62 所示的对话框。

4) 在图 45.6.62 所示对话框中, 单击 [OK], 得到如图 45.6.63 所示对话框, 选用 [Variable], 得到图 45.6.64 所示的对话框。

5) 选用分析变量是 1-X。单击 [OK], 得到如图 45.6.65 所示对话框。

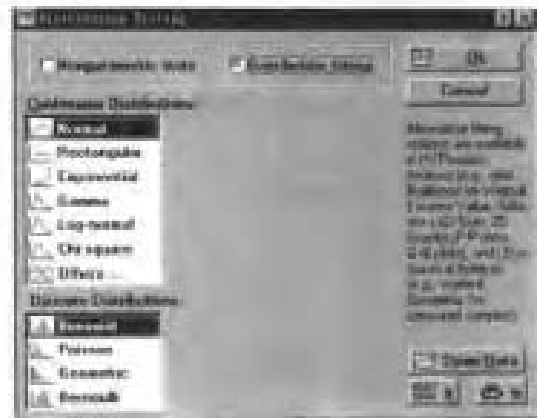


图 45.6.62 Distribution Fitting (分布拟合) 对话框

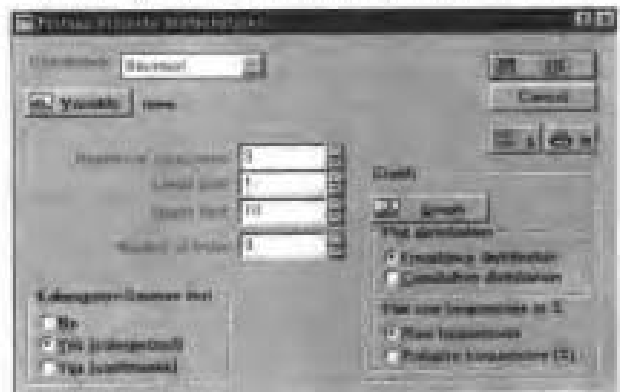


图 45.6.63 Fitting Discrete Distributions (拟合离散分布) 对话框

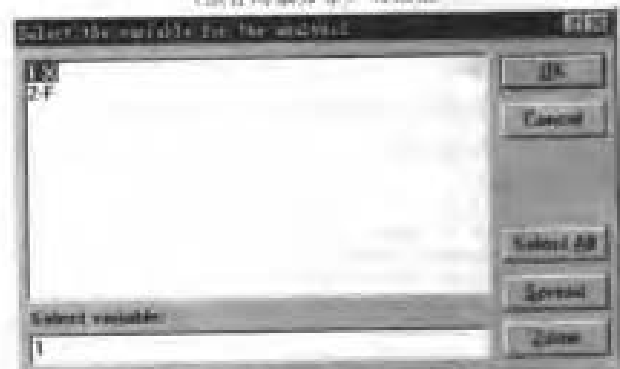


图 45.6.64 选择变量 (Select variable)

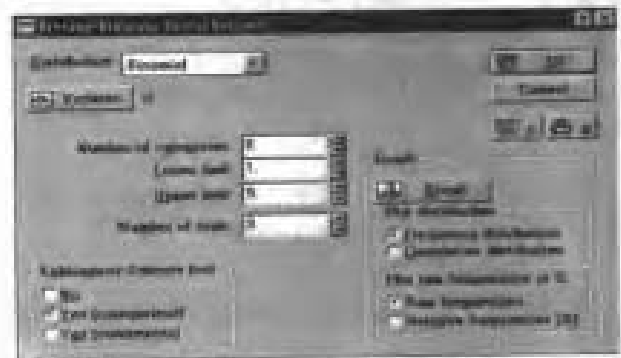


图 45.6.65 已选拟合二项分布 (Binomial)

6) 在图 45.6.65 所示对话框中, 单击 [OK], 得到计算结果。

Upper Boundary	observed	cumulative	percent	cumul. %	expected	cumulative
上边界	观察频数	累计观察数	观察百分数	累计观察百分数	期望频数	累计期望数
<=1	0	0	0.00000	0.0000	1.6073	1.6073
2-	2	2	4.00000	4.0000	1.9342	1.09495
3-	2	4	4.00000	8.0000	3.3944	4.3857
4-	9	13	18.00000	26.0000	7.92489	12.3106
5-	12	25	24.00000	50.0000	12.34762	24.73408
6-	12	37	24.00000	74.0000	12.80196	37.55604
7-	9	46	18.00000	92.0000	6.54357	44.09961
8-	3	49	6.00000	98.0000	3.32520	47.42481
Infinity	1	50	2.00000	100.0000	1.67519	50.00000

Upper Boundary	percent	cumul. %	observed
上边界	期望百分数	累计期望百分数	观察频数
<=1	32.146	32.15	1.6073
2-	1.86844	2.1890	1.06578
3-	6.78725	8.9771	1.39366
4-	15.84978	24.8269	1.07541
5-	24.07524	49.5022	0.33762
6-	25.60951	75.1121	1.80496
7-	17.08714	92.1992	1.45643
8-	6.65049	98.8497	1.32520
Infinity	1.15623	100.0000	1.42481

### 图形结果

单击 [Graph], 得到如图 45.6-60 所示的图形。由于  $p(p=0.9564728) > 0.95$ , 可以认为服从二项分布, 拟合优度较好。

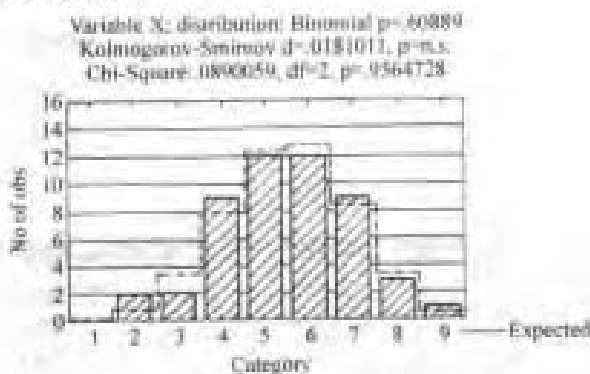


图 45.6-60 二项分布图 (期望二项分布曲线与直方图)

### 2. 泊松分布 (Poisson)

泊松分布 (Poisson) 是一种很重要的离散型分布, 常用于研究单位时间内 (或单位空间内) 某事件发生次数的分布情况。

例 已知 300 个单位时间内的试件失效数结果如表 45.6-11。问此资料是否服从 Poisson 分布。

1) 建立数据文件: poisson2.sta (见图 45.6-67)。由于本资料是频数表, 应先对频数 (F) 加权, 然后, 再对单位时间内的试件失效数 (X) 进行拟合。

表 45.6-11 300 个单位时间内的试件失效数

300 个单位时间内的试件失效数 (X)	0	1	2	3	4	5	6	7
观察频数 (F)	26	51	84	70	42	15	9	3

图 45.6-67 数据文件

2) 单击 [Options] ⇒ [Weight], 得到如图 45.6-68 所示界面。权重变量 (Weight variable) 是 F, 状态变量 (Status) 是 ON。单击 [OK] (完成对变量 F 的加权)。

3) 单击 [Nonparametrics/Distrib.] (非参数统计检验/分布拟合) ⇒ [Switches] [Distribution fitting] (分布拟合) ⇒ [Poisson] (泊松分布), 得到如图 45.6-



图 45-6-68 定义权重变量

69 所示界面。

4) 在图 45-6-69 所示界面中, 单击 [OK], 得到如图 45-6-70 所示界面。在图 45-6-70 所示界面中, 单击 [Variable] (待选变量), 得到如图 45-6-71 所示界面。

5) 在选择变量 (Select variable) 界面中, 选用 [2-X] → [OK], 得到如图 45-6-72 所示对话框。

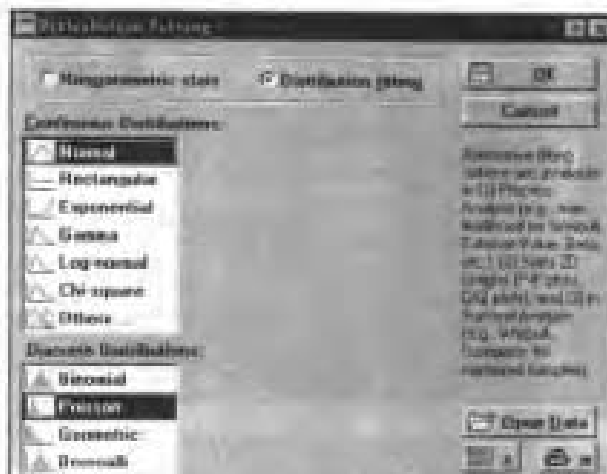


图 45-6-69 分布拟合

6) 在图 45-6-72 所示对话框中, 单击 [OK], 得到计算结果如下。

Upper Boundary	observed	cumulative	percent	cumul. %	expected	cumulative
上边界	观察频数	累计观察数	观察百分数	累计观察百分数	期望频数	累计期望数
0	28	28	8.6667	8.667	24.87300	24.8730
1	31	77	12.0000	20.667	63.53978	80.6098
2	84	161	28.0000	53.667	77.10751	163.9148
3	70	231	23.3333	77.0000	68.99223	227.2135
4	42	273	14.0000	91.0000	39.83952	287.7530
5	15	288	5.0000	96.0000	18.84028	287.5931
6	8	297	2.0000	98.0000	8.23383	295.8267
Infinity	3	300	1.0000	100.0000	4.17327	300.0000

Upper Boundary	percent	cumul. %	observed
上边界	期望百分数	累计期望百分数	观察频数
0	8.29100	8.2910	1.1270
1	20.64405	28.9350	10.3338
2	25.70231	54.6381	5.8925
3	21.33308	75.9712	6.0008
4	13.27984	89.2510	2.1603
5	6.81236	96.0634	4.8401
6	2.74454	98.8080	1.7604
Infinity	1.29100	100.0000	1.1733

**图形结果**

单击 [Graph], 得到如图 45-6-73 所示图形。可知  $P(p=0.4759851) > 0.15$ , 故可认为本资料服从泊松

分布 (Poisson)。

类似的方法, 可以做其他离散型分布的拟合。

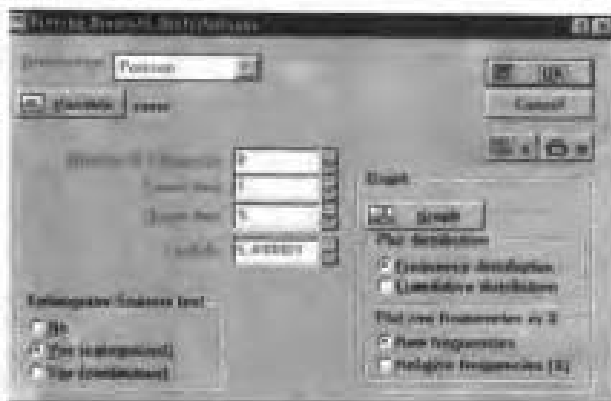


图 45-6-70 拟合离散分布



图 45-6-71 选择变量

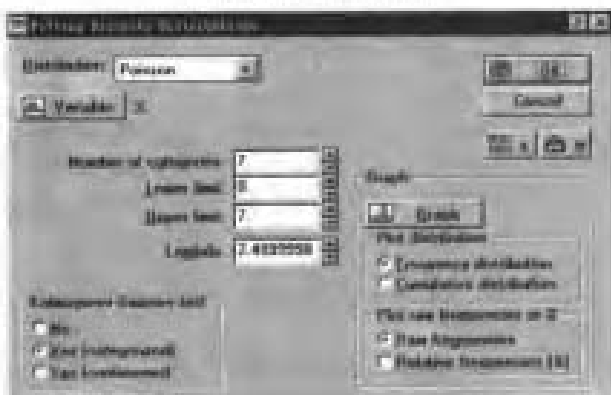


图 45-6-72 泊松拟合分布

Variable X: distribution, Poisson  $\lambda=2.4000$   
 Kolmogorov-Smirnov  $d=0.126892$ ,  $p=0.1$ ,  
 Chi-Square: 4.529315,  $df=5$ ,  $p=0.4759851$

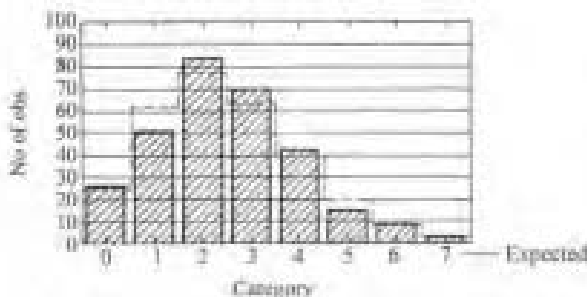


图 45-6-73 拟合分布图(期望泊松分布曲线与直方图)

### 6.4.3 概率坐标纸

概率坐标纸是一种特殊的坐标纸,在进行可靠性分析时,经常要用到它,应用概率坐标纸即可以判断被抽样的母体是否符合相应的分布,又可以通过它获得分布参数估计值,其检验方法为,将累计概率数据在概率坐标纸上描点,若各点明显不在一直线上,则认为该随机变量服从该分布的假设不成立,分析停止,另作其他假设;如果各点基本在一直线上,则接受服从分布的假设,并从坐标纸上获取取分布参数估计值。

Statistics for Windows 的二维统计图中 (Stats 2D Graphs), 可以作直方图 (Histograms)、分位数-分位数图 (Quantile-Quantile Plots)、概率-概率图 (Probability-Probability Plots)。

#### (1) 概率-概率图 (Probability-Probability Plots)

概率-概率图 (Probability-Probability Plots) 具有按制定的分布类型进行拟合的功能。

例 对 10 辆使用时间相同的车辆的制动闸瓦进行磨损量测定,其磨损量数据为:0.204, 0.231, 0.157, 0.191, 0.261, 0.127, 0.217, 0.173, 0.244, 0.291, 应用概率坐标纸判断其是否符合正态分布类型。

1) 建立数据文件: probability.sts (见图 45-6-74)。

DATA	VAR1
1	191
2	204
3	231
4	157
5	261
6	127
7	173
8	244
9	291
10	717

图 45-6-74 数据文件

2) 单击 [Graphs] (图形), 得到如图 45-6-75 所示界面。

3) 单击 [Stats 2D Graphs] (二位图形), 得到如图 45-6-76 所示界面。

4) 单击 [Probability-Probability Plots] (概率-概率图), 得到如图 45-6-77 所示对话框, 其中包括:

- Variable 在变量界面中, 选用 [VAR1]。
- Distribution 中可以选择如下分布函数:
  - ⊙ Beta, 贝塔 ( $\beta$ ) 分布。
  - ⊙ Exponential, 指数分布。
  - ⊙ Extreme, 极端值。



图 45-6-75 [Graphs] 界面

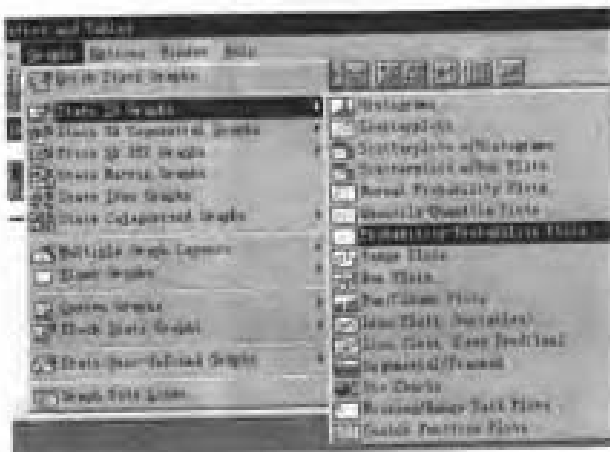


图 45-6-76 [Stats 2D Graphs] (二维图形) 界面



图 45-6-77 [Probability-Probability Plots] (概率-概率图) 对话框

- ☆ Gamma,  $\Gamma$  分布,
- ☆ Lognormal, 对数正态分布,
- ☆ Normal, 正态分布,
- ☆ Rayleigh, 瑞利分布,
- ☆ Weibull, 威布尔分布,

5) 在图 45-6-77 所示对话框中,选择 [Normal] 分布,单击 [OK],得到如图 45-6-78 计算结果。

**图形结果**

1) 从图 45-6-78 可以目测到,正态坐标纸上实验数据基本在一直线,可以接受,可以接受车辆的制动闸瓦进行透损量符合正态分布的假设。

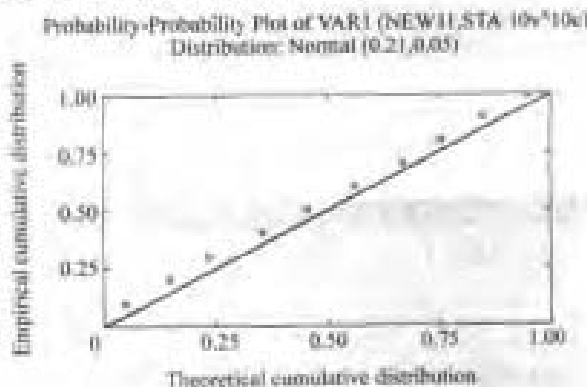


图 45-6-78 概率-概率图图形

2) 图中还给出了采用极大似然法估算的均值和方差, Distribution: Normal (0.21, 0.05)。

**(2) 分位数-分位数图 (Quantile-Quantile Plots)**

分位数-分位数图 (Quantile-Quantile Plots) 具有按指定的概率分布函数绘制坐标值,并给出各分析变量所对应的实际累计概率分布(分位数)拟合直线。以两参数威布尔分布为例予以介绍。

例 为查明某产品的寿命分布情况,抽查 6 个试样实验得 6 个寿命值为:50800, 98000, 161000, 215000, 304000, 412000。假设寿命服从威布尔分布,用概率坐标纸法检验其分布假设。

其分析方法与步骤与概率-概率图 (Probability-Probability Plots) 类似:

- 1) 建立数据文件。
- 2) 单击 [Graphs] (图形),得到如图 45-6-75 所示界面。
- 3) 单击 [Stats 2D Graphs] (二维图形),得到如图 45-6-76 所示界面。
- 4) 单击 [Quantile-Quantile Plots] (分位数-分位数图),得到如图 45-6-80 所示对话框,其中包括:
  - Variable 在变量界面中,选用 [VAR2]。
  - Distribution 中可以選擇如下分布函数:
    - ☆ Beta, 贝塔 ( $\beta$ ) 分布,



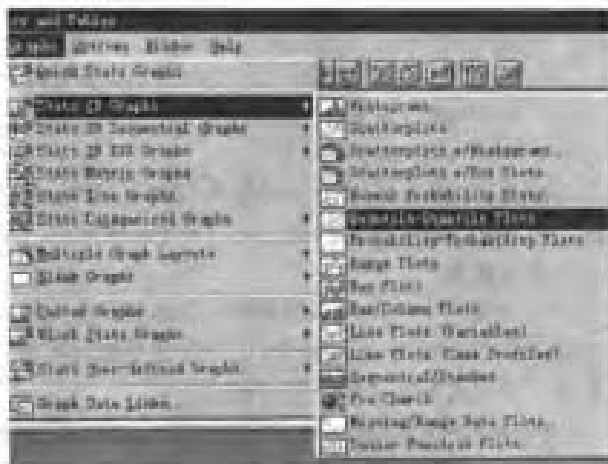


图 45-6-79 [Stat>2D Graphs] (二位图形) 菜单 [Quantile-Quantile Plot]

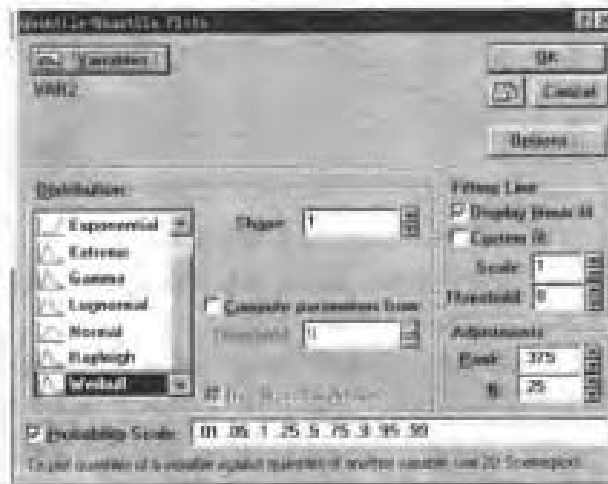


图 45-6-80 [Quantile-Quantile Plots] (分位数-分位数数据) 对话框

- ☆ Exponential, 指数分布。
- ☆ Extreme, 极值。
- ☆ Gamma,  $\Gamma$  分布。
- ☆ Lognormal, 对数正态分布。
- ☆ Normal, 正态分布。
- ☆ Rayleigh, 瑞利分布。
- ☆ Weibull, 威布尔分布。

● Probability Scale (概率比例尺), 累积概率值。

本例中采用默认值, 0.01, 0.05, 0.10, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99。

5) 在图 45-6-80 所示对话框中, 选择 [Weibull] 分布, 单击 [OK], 得到如图 45-6-81 计算结果。

#### 图形结果

1) 从图 45-6-81 可以目测到, 威布尔坐标纸上实验数据基本在一直线, 可以接受其产品的寿命服从威

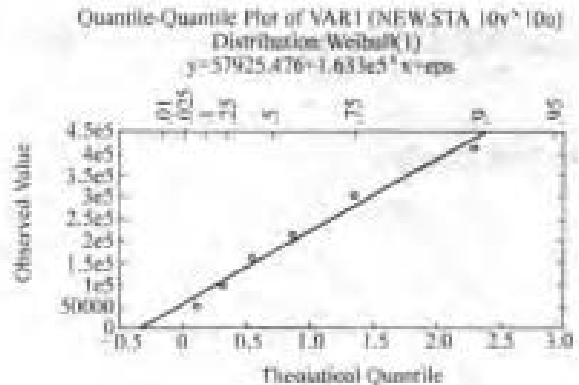


图 45-6-81 分位数-分位数数据图形

布尔分布的假设。

2) 进行参数的点估计: 威布尔坐标纸上实验数据基本在一直线位置参数,  $\hat{a} = 0$ ; 尺度参数,  $P(t) = F(N) = 63.2\%$  的水平线所绘制直线校点的横坐标就是尺度参数, 即  $\hat{b} = N_{63.2\%}$ 。本例从图 45-6-81 中可读取  $\hat{b} = 24000$ 。

### 主要符号表

$a$	威布尔分布的位置参数
$A$	有效性, 有效度
$A(t)$	有效度
$b$	威布尔分布的尺度参数
$c$	循环, 次数
$C$	置信度, 费用, 刚度系数
$C_r$	致命度系数
$D_{k,s}$	K-S 检验临界值
ERT	修复时间分布的中位数
$E(x)$	随机变量 $X$ 的均值
$F$	失效概率
$f(x)$	随机变量 $X$ 的概率密度函数
$F(x)$	随机变量 $x$ 的累积分布函数 (简称分布函数)
$I_0$	概率重要度
$k$	威布尔分布的形状参数, 分组数
$k_r$	有效应力集中系数
$M$	维修性, 维修度
$m(t)$	维修时间的概率密度函数
$M(t)$	维修度
MTBF	平均无故障工作时间
MTTF	失效前平均工作时间
MTTR	平均修复时间
$n$	工作安全系数, 工作循环次数, 子样容量
$n_r$	有效子样容量
$N$	失效循环次数, 疲劳循环数
$P$	概率, 显著性检验值

$P(A)$	事件 $A$ 发生的概率	$\varphi(u)$	标准正态分布的概率密度函数
$P(A B)$	在事件 $B$ 已经发生的条件下事件 $A$ 发生的条件概率	$\Phi(u)$	标准正态分布函数

## 参 考 文 献

- |                |                   |    |  |
|----------------|-------------------|----|--|
| $q$            | 敏性系数              | 1  | 王少萍. 工程可靠性[M]. 北京:北京航空航天大学出版社,2000   |
| $r$            | 应力比               | 2  | 黄祥瑞. 可靠性工程[M]. 北京:清华大学出版社,1989   |
| $R$            | 可靠性,可靠度           | 3  | 王世萍,朱敏波. 电子机械可靠性与维修性[M]. 北京:清华大学出版社,2000                                   |
| $R_{L1}$       | 可靠度的单侧置信区间下限      | 4  | 刘维信. 机械可靠性设计[M]. 北京:清华大学出版社,1996   |
| $R_{L2}$       | 可靠度的双侧置信区间下限      | 5  | 朱继洲,奚树人等. 核反应堆安全分析[M]. 西安:西安交通大学出版社,2000                                   |
| $R_{U2}$       | 可靠度的双侧置信区间上限      | 6  | 何国伟. 可信性工程[M]. 北京:中国标准出版社,1997   |
| $R_{U1}$       | 可靠度的单侧置信区间上限      | 7  | 张义民. 汽车零部件可靠性设计[M]. 北京:北京理工大学出版社,2000                                      |
| $s$            | 子样标准差             | 8  | 朱敏波,王世萍,韩芳明. 计算机辅助故障树的建立与分析[J]. 计算机辅助工程,1998,(1):35~38                     |
| $s_x$          | 随机变量 $X$ 的子样标准差   | 9  | 桑怀胜,王浩,郁文贤. 动态建造故障树的计算机实现[J]. 现代计算机,1999,(2):49~52                         |
| $s^2$          | 子样方差              | 10 | 王朝辉. 复杂系统故障树的计算自建原理及其技术[J]. 第四届全国机械设备故障诊断学术会议论文集,1994                      |
| $t$            | 时间                | 11 | 简志敏,胡东成. 计算机辅助建造故障树的算法[J]. 第三届全国计算机应学术交流大会论文集(下),北京:电子工业出版社,1995,1666~1669 |
| $T$            | 转矩,扭矩             | 12 | 张祖明. 计算机辅助机械可靠性设计[J]. 第三届全国计算机应学术交流大会论文集,北京:电子工业出版社,1995,229~232           |
| $T_p$          | 时间间隔              | 13 | 夏家海,等. 基于立方体理论 MFTA 软件设计与实现[J]. 中国控制与决策学术年会论文集,2000,422~426                |
| $u_p$          | 标准正态偏量            | 14 | 徐灏. 机械强度的可靠性设计[M]. 北京:机械工业出版社,1984   |
| $V(x)$         | 随机变量 $X$ 的方差      | 15 | 王锡吉. 可靠性工程技术[M]. 北京:电子工业出版社,1995   |
| $\bar{x}$      | 随机变量 $X$ 的子样均值    | 16 | 杨志飞,李平. 以故障中心的工程可靠性[M]. 北京:国防工业出版社,1995                                    |
| $z$            | 可靠度系数,联接系数        | 17 | 王熙珍. 可靠性. 冗余及容错技术[M]. 北京:航空工业出版社,1991                                      |
| $\alpha$       | 风险,显著性水平          | 18 | 梅启智,廖炯生,孙惠中. 系统可靠性工程基础[M]. 北京:科学出版社,1992                                   |
| $\alpha_s$     | 理论应力集中系数          | 19 | 陆延孝,郑鹏洲. 可靠性设计与分析[M]. 北京:国防工业出版社,1995                                      |
| $\beta$        | 表面加工系数            |    |  |
| $\gamma$       | 风险,置信度的风险         |    |  |
| $\epsilon$     | 尺寸系数              |    |  |
| $\lambda$      | 失效率               |    |  |
| $\mu$          | 母体均值              |    |  |
| $\mu(t)$       | 修复率               |    |  |
| $\nu$          | 自由度,变异系数,泊松比      |    |  |
| $\rho$         | 相关系数              |    |  |
| $\sigma$       | 应力,母体标准差          |    |  |
| $\hat{\sigma}$ | 母体标准差估计值          |    |  |
| $\sigma_s$     | 应力幅               |    |  |
| $\sigma_b$     | 抗拉强度              |    |  |
| $\sigma_c$     | 当量应力              |    |  |
| $\sigma_m$     | 平均应力              |    |  |
| $\sigma_q$     | 挤压应力              |    |  |
| $\sigma_s$     | 屈服点               |    |  |
| $\sigma_r$     | 应力比为 $r$ 的疲劳极限    |    |  |
| $\sigma_{-1}$  | 应力比为 $-1$ 的疲劳极限   |    |  |
| $\theta$       | 平均寿命              |    |  |
| $\tau$         | 切应力               |    |  |
| $\tau_s$       | 切应力幅              |    |  |
| $\tau_m$       | 平均切应力             |    |  |
| $\tau_{-1}$    | 应力比为 $-1$ 的剪切疲劳极限 |    |  |

- 20 徐维新,秦英孝. 可靠性工程[M]. 北京:电子工业出版社,1988
- 21 刘品. 可靠性工程基础[M]. 北京:中国计量出版社,1995
- 22 浦维达. 汽车可靠性工程[M]. 北京:机械工业出版社,1998
- 23 周广淘. 计算机辅助可靠性工程[M]. 北京:宇航出版社,1990
- 24 俞全余. 计算机辅助求解减速器故障树的最小割集[J]. 机械设计与研究,1996,(2):44~45
- 25 丁彩虹,黄文虎,姜兴渭. 基于 Petri 网的故障树分析方法[J]. 强度与环境,1999,(4):31~37
- 26 郑裕国,张康达. 故障树定性和定量分析的算法[J]. 浙江工业大学学报,1995,23(1):42~47
- 27 季细星. 游梁式抽油机的故障树分析[J]. 石油机械,1994,22(2):41~47
- 28 刘永宾,陈金水,谢学武. 割集矩阵在早期不交化的 FTA 中的应用[J]. 天津大学学报,2000,33(3):31~36
- 29 宋笔锋,王小,张怡哲. 失效树分析的矩阵化方法[J]. 机械科学与技术,1997,16(6):
- 30 陆愈实,扬静. 故障树可视化分析系统[J]. 中国安全科学学报,2000,10:80~84
- 31 闫剑平. 一种故障树模块的划分方法[J]. 北方交通大学学报,2000,24(5):
- 32 孙晓,卢新田等. 风机系统可靠性分析[J]. 吉林工业大学学报,1995,25(2):94~99
- 33 Henley E J, Kumamoto H. Reliability Engineering and Risk Assessment[M], 1981
- 34 Barlow R E, et al. Reliability and Fault Tree Analysis[J]. SIMA, 1975
- 35 Kiran Kumar Vemuri. Automatic Synthesis of Fault Trees for Computer-Based Systems [J]. IEEE Transactions on Reliability, 1999, 48(4): 394~402
- 36 Lapp S A, Powers G J. Computer-aided synthesis of fault trees[J]. IEEE Trans. Reliability, 1977, 26:2~13
- 37 Kocza G, Bossche A. Automatic fault-tree synthesis and real-time tree trimming based on computer models. Proc [J]. Annu. Reliability & Maintainability Symp, 1997, (3):71~75
- 38 Dugan J B, Bavuso S, Boyd M. Dynamic fault-tree models for fault tolerant computer systems [J]. IEEE Trans. Reliability, 1992, (41):363~377
- 39 Powers G J Lapp, SA. Fault-tree models for the analysis of complex computer-based systems. Proc [J]. Annu. Reliability & Maintainability Symp, 1996, 200~207
- 40 Xie Liyang. A Knowledge based multi-dimension discrete common cause failure model[J]. Nuclear Engineering and Design, 1998, 107~116
- 41 Ren Yangsong, Joanne Bechta Duyan. Design of Reliable Systems Using Static & Dynamic Fault Tree [J]. IEEE Trans. Reliability, 1998, 47(3): 234~244
- 42 Mosteh Ali. Common Cause Failures: An Analysis Methodology and Examples [J]. Reliability Engineering System Safety, 1991, 34:249~292
- 43 Rasmuson Dale M. Some Practical Considerations in Treating Dependencies in PRAs [J]. Reliability Engineering & System Safety, 1991, 34:327~343
- 44 朱文予. 机械概率设计与模糊设计[M]. 北京:高等教育出版社,2001
- 45 肖岗,李天柁. 系统可靠性分析中的蒙特卡罗方法[M]. 北京:科学出版社,2003
- 46 吴波,黎明发. 机械零件与系统可靠性模型[M]. 北京:化学工业出版社,2003
- 47 王超,王金. 机械可靠性工程. 北京:冶金工业出版社,1992
- 48 徐灏主编. 机械设计手册,第13篇可靠性设计(王超主编). 北京:机械工业出版社,1991
- 49 第四机械工业部标准化研究所. 可靠性试验用表. 增订本. 北京:国防工业出版社,1987
- 50 戴树森,费鹤良,王玲玲,苏德清,白鹤翔,滕怀流. 可靠性试验及其统计分析上下册. 北京:国防工业出版社,1984
- 51 机械工程材料性能数据手册编委会. 机械工程材料性能数据手册. 北京:机械工业出版社,1995
- 52 牟致忠. 机械零件可靠性设计. 北京:机械工业出版社,1988
- 53 牟致忠. 维修性设计. 上海:上海科学技术文献出版社,1990
- 54 卢玉明. 机械零件的可靠性设计. 北京:高等教育出版社,1989
- 55 高镇同. 疲劳应用统计学. 北京:国防工业出版社,1986
- 56 徐维新,秦英孝. 可靠性工程. 北京:电子工业出版社,1988
- 57 肖德辉. 可靠性工程. 北京:宇航出版社,1985
- 58 曾天翔等译. 可靠性设计手册第1卷(译自美国

- MIL-HDBK-338 电子设备可靠性设计手册, 1976 版). 北京: 航空工业出版社, 1987
- 59 张训诰, 肖德辉. 可靠性及其应用. 北京: 兵器工业出版社, 1991
- 60 顾履平, 冯锡曙. 实用可靠性技术. 北京: 机械工业出版社, 1992
- 61 何国伟. 可靠性设计. 北京: 机械工业出版社, 1993
- 62 徐灏. 机械强度的可靠性设计. 北京: 机械工业出版社, 1984
- 63 徐灏. 疲劳强度. 北京: 高等教育出版社, 1988
- 64 徐灏. 概率疲劳. 沈阳: 东北大学出版社, 1994
- 65 徐灏, 胡俏. 随机载荷下疲劳可靠性计算模型. 北京: 机械工程学报(英文版), 1994, 7(4);
- 66 徐灏, 胡俏. 随机载荷下零件的可靠度计算新方法. 北京: 农业机械学报, 1995, 26(2);
- 67 徐灏, 王维明, 吴志学. 短裂纹的裂纹尺寸分布密度函数与耐久性设计. 北京: 航空学报, 1995, 16(5);
- 68 洪楠主编. STATISTICA for Windows 统计与图表分析教程. 北京: 清华大学出版社, 北方交通大学出版社, 2002

# 第 46 篇 实验设计法

主 编 林忠钦  
编写人 金 伟 李余兵

# 第 1 章 实验设计与正交实验表

## 1 实验设计的作用

### 1.1 实验设计的概念

实验设计是通过一系列实验进行规划的方法,其目的是利用统计方法为一种或多种产品或工艺参数确定最佳环境。实验设计在要求产品和工艺质量有所突破时使用,它帮助开发者搞清楚材料或某工艺参数的变化对产品或工艺质量产生什么影响。

田口玄一 是实验设计法的创始人,本篇主要参考了《实验设计法》的译著和原著。

### 1.2 实验设计的领域

领域主要有机械、化工、医学、制药等。对于一个或多个目标变量,有下述三种情况:

1) 在函数关系  $y=f(A, B, \dots)$  式中,在变数  $A, B, \dots$  变化的范围内,函数形式  $f$  处于未知或不太了解的情况下,在研究的范围内决定变数  $A, B, \dots$  的水平,可以通过实验与试制用公式与曲线表示出函数关系。变数  $A, B, \dots$  的水平不能独立的变化或者目标变量是计量值与计数分类值等混合在一起的情况是常见的,如上所述,灵巧的实验设计技术是有效的手段。

2) 函数关系式  $y=f(A, B, \dots)$  中,关于函数的形态是有假定与预想的,由于其中存在有未知参数,所以希望对那些未知参数进行估计。这里利用二乘法等方法,根据对应于  $A, B, \dots$  的各种观测值  $y$ ,对未知参数作出估计。预测未知参数存在范围的知识也应该考虑,这类型的问题应该结合理工学、社会科学的数据,建立理论式与公式加以表现,它在生产工序等数据分析中也经常出现。

3) 函数关系  $y=f(A, B, \dots)$  为已知的情况,这是在数值计算、设计计算、规划计算等计算中经常出现的问题。如电路设计方面,为求如何布置  $L, R, C$  等电路常数,输出  $y$  才达到目标值的问题,必须将电路常数做各种的变化,解联立微分方程式,调查研究对  $y$  值的影响。当电路常数的变数比较多时,所有组合都计算是很困难的,有效的办法是利用正交表的差分计算方法。

实验设计的应用领域是处理多变数的函数,利用正交表作为一部分组合的实验与计算的所谓配列问题,与应用方差分析评价因素效应大小的解析问题。

实验设计在工业生产和工程设计中能发挥重要的作用,例如:

- 1) 提高产量;
- 2) 减少质量的波动,提高产品质量;
- 3) 大大缩短新产品试验周期;
- 4) 降低成本;
- 5) 实验设计延长产品寿命。

一次合理的实验安排应做到:

- 1) 实验次数尽可能少;
- 2) 便于分析和处理实验数据;
- 3) 能得到满意的结果。

### 1.3 实验顺序的随机性

实验的随机化是费歇 (R. A. Fisher) 开发的技术方法。在实验数据分析当中,对这一问题必须始终考虑为假定。在实验中,若人为地有次序地安排实验会产生系统误差,从而混淆了因素对效应有无作用的判断;而一旦有了系统误差,就不能通过任何数据处理方法来消除,有时会使实验作出不正确的判断而导致失败。为了消除系统误差,在安排实验时对各种排列采用随机化的方法是有效的。

但是根据实验情况,也不一定非用这种方法不可,即使在本实施实验随机化的情况下,也必须认为实施了随机化,在此条件下进行数据分析。

所谓随机化,就是在实验中对实验顺序和因素水平的排列的顺序按照随机性原则来安排。这一原则的执行,可以正确地估计实验误差,减少误差,提高试验的可靠性和再现性。随机性原则的实施一般可以借助于随机数表来安排实验。

## 2 实验设计中因素和水平的确定方法

当进行实验与调查时,对实验结果特性值(指标)可能有影响的要考虑的各种各样原因或要素即为因素,或称为因子。通常是在最适当条件下的各种变素范围内,进行各种变化的实验。

关于因素的数量,虽然根据实验的不同而有所不同,但是从更好地利用实验设计法的角度来看,至少也要 7~8 个,一般不论多少都可以,除此之外,希望实验中取用的因素尽可能的具体。

实验设计中选定的因素所处的状态和条件的变化,可能引起指标的变化,称因素的变化状态和条件

叫做水平或者位级。

## 2.1 因素的分类

最简单的分类把因素分为可控因素和不可控因素。加热温度、融化温度、切削速度、进给量等人们可以控制和调节的因素称为可控因素；机床的微振动、刀具的磨损等人们暂时不能控制和调节的因素称为不可控因素。实验设计一般仅适用于可控因素。

按因素的作用，因素一般可以分为五类：

### (1) 控制因素

为了选择最宜的水平，提出若干个水平在实验中取用的因素为控制因素。控制因素的水平可以进行比较而且容易人为选择。例如机械加工中的切削速度、进给量、背吃刀量；电子产品的电容值、电阻值；化工产品的温度、压力、催化剂种类。一般说的因素即为控制因素，没有控制因素的实验叫做调查。

### (2) 标示因素

标示因素是指外界的环境条件、产品的使用条件等的因素。这种因素虽然和控制因素同样具有水平，但对选择最宜的水平是没有意义的。例如在日光灯的制选实验中，设定研究寿命问题，虽然只对标准电压 100V 做实验。但还取 80V、120V 等电压做实验，调查在 80V、100V、120V 的寿命。构成标示因素的有下列情况：

- 1) 物品的各种各样使用条件；
- 2) 实验条件（强制老化等情况）；
- 3) 老化过程的时间；
- 4) 品种；
- 5) 人、装置等差异。

### (3) 区组因素

区组因素是具有水平，但其水平没有技术意义的因素。因此这种因素不能起调节的作用，是为了不使其水平混淆到控制因素中，而采用这种因素。地区之差、日期差、批量差、操作时间之差等，根据取用这种方法，也有变成标示因素的，但一般均为区组因素。

### (4) 辅助因素

辅助因素包括上述三种因素所影响的辅助实验记录，记录随机实验条件状态下的辅助测试值。前者为中间实验值，后者为配列上的随机的标示因素或区组因素。

### (5) 误差因素

不测试其状态，对实验值有影响无法控制的全部其他因素为误差因素，简称误差，也就是影响产品质量的外干扰、内干扰、随机干扰的总和。根据实验的随机化方法，划分为一次误差、二次误差等阶段。

## 2.2 因素间相互关系

因素多的时候，考虑因素间的相互关系，对于决定

这些因素的水平是非常重要的。

应尽可能多地取用对实验目的起作用的控制因素与标示因素，这意味着相关人员的技术经验丰富，观察力强。取用效应大的区组因素与辅助因素，表明相互人员的经验与知识丰富。如果无用的因素取用得多，则表明知识不足，因为不管控制因素多少，必要的实验规模都变化不大，所以与其注意担心取用无用因素不如注意防止必要的因素的遗漏。

分析相互关系前，需先了解主效应和交互作用。主效应就是在其他因素的各种水平中，涉及到该因素的实验值的效应的平均。

提高研究效率的最好方法是利用小规模实验得出大规模生产中应用的结果。设有两类材料  $A_1$ 、 $A_2$ ，在改变各种条件（温度  $B$ 、搅拌条件  $C$ 、压力  $D$ 、掺杂剂  $E$  等）时， $A_1$ 、 $A_2$  的差见表 46-1-1。

表 46-1-1 不同材料在不同条件下的差值

	$B_1$	$B_2$	$C_1$	$C_2$	$D_1$	$D_2$	$E_1$	$E_2$
$A_1$	60	50	72	70	58	68	69	72
$A_2$	68	61	80	76	64	75	73	79
差值	+8	+6	+8	+6	+5	+7	+4	+7

如果已知  $A_2$  比  $A_1$  好，其差值是在  $6.5 \pm 3.0$  范围内，同样，当大规模生产时， $A_2$  良好的可能性就大。如果  $A_2$ 、 $A_1$  之差在各种条件下结果不完全相同，那么就或多或少存在交互作用。对主效应与交互作用进行比较，只有当交互作用越小时，结果才越可靠，所以尽量减小交互作用。控制因素的交互作用变大，就意味着实验结果的可靠性显著下降。交互作用的有与无从因素的性质来看是由于特性值缺乏单调性的缘故，因此只要抓住具有单调性的特性值实验结果就可靠了。

在特性值  $y$  为

$$y \propto A^{\alpha} B^{\beta} C^{\gamma}$$

的情况下，三因素具有交互作用，取  $y$  的对数

$$\lg y = \alpha \lg A + \beta \lg B + \gamma \lg C + \text{const}$$

则不存在交互作用。控制因素间的交互作用倘若不能够变小，实验则达不到高效化。

在决定因素的水平时，将实验的因素按照工艺顺序排列，由开始的因素决定其水平，这些因素设定为  $A$ 、 $B$ 、 $C$ 、...，现决定因素  $A$  的实验范围，如果是三水平，一般将其范围的下端、中端、上端定为  $A_1$ 、 $A_2$ 、 $A_3$ 。同样依次定义四水平、五水平。对于二水平的情况，现在的水平为  $A_1$ ，新的方案的水平为  $A_2$ 。此时，实验前已知明显坏的水平（考虑成本的基础上），不要包括在  $A$  的实验范围内，然而欲取的实验范围应全部包括在内。其次决定  $B$  的水平必须考虑  $A$  的水平不同， $B$  欲取的实验范围是否不同的问题。尽可能欲取的实验范围全部都做，已知实际不使用的范围坚决不做

实验。实际情况中,实验人员往往为了不出大的差错,将已知不佳的因素变化范围也都做实验。实验必须限于在不能判断的范围内取得信息。

### 2.3 多因素情况的处理

很多因素间相互关联时,将考虑相互关系的水平全部写出来的工作量很大。最适宜组合被排除在实验范围之外的损失一般要比做坏实验的组的损失要大,必须考虑相互关系时,最好是单独把因素水平标出来。标示法的优点就是不能因为其他因素的水平值的改变而改变水平的范围。三水平标示法如下:

$$\begin{cases} A_1 & -10\% \\ A_2 & \text{理论值} \\ A_3 & +10\% \end{cases} \begin{cases} B_1 & -5\% \\ B_2 & \text{标准方法求值} \\ B_3 & +5\% \end{cases}$$

$$\begin{cases} C_1 & -2\% \\ C_2 & \text{现在决定的方法求值} \\ C_3 & +2\% \end{cases}$$

以上水平也可类似处理。可以相应列表标示。

考虑到各因素间的交互作用,定出因素的水平,避免不必要范围的实验,这样不仅具有只进行必要范围实验的优点,而且可以消去交互作用。如果控制因素的水平按照下述方法解决,可以较合理地决定水平:

- 1) 对于那些要做实验的因素,即认为最适宜的一切组合,都要选出,要实验的组合一个也不能漏掉。
- 2) 定出某种组合后,在其结果实验前已明确明显的不好或者实验上不适用的组合,一个也不要包括在内。

### 2.4 水平的处理

根据分类定水平,例如为了调查组装产品与系统的特性,受元件与结构因素的影响,可以找出元件与因素的制造条件的研究方法。从人类看,有以下三种情况:

- 1) 分解多少次都可以再组装的情况;
- 2) 可以分解一次再组装,且不合格与损坏的情况少;
- 3) 不能分解再组装的情况。

在选取水平时应注意以下几点:

#### (1) 水平宜选三水平

因为三水平的因素实验结果分析的效应图分布多数是二次函数曲线,而二次曲线有利于观察实验趋势,这对分析试验有利。

应当充分发挥专业技术作用所取定的因素水平,尽可能取在最佳区域中或者接近最佳区域,按选择的因素水平做实验,其效率会高些;当专业技术较低时,可能因素水平取不到最佳区域附近。这时可以把水平

间隔拉开,尽可能使最佳区域包含在水平区间内。然后通过1~2次实验逐渐缩小水平区间,求出最佳状态或条件,不满意时可重复试验。

#### (2) 水平取等间隔原则

水平的间隔是由技术水平、技术知识范围所决定的。水平的等间隔一般是取算术等间隔值,在一些情况下可以取对数等间隔。由于各种客观条件的限制和技术上的原因,在取等间隔值时会有差值,但尽可能使差值缩小些,一般不超过20%的间隔值。

#### (3) 选取的水平应是具体的

水平是具体的,指的是可以直接控制的,并且水平的变化要能直接影响试验的指标,使其有不同程度的变化。

因素的水平通常用1、2、3...表示。

### 2.5 各种辅助方法

对于不能自由选择水平的因素,可以采取下列方法:①分层法;②辅助测试;③随机化。

分层法本质上就是分类问题。事先可以做分类的情况下,必须取为标示因素或者区组因素。

## 3 正交实验表

在化工工艺开发研究中,通常以反应的收率或选择性作为优化目标。反应温度、压力、原料的摩尔比、反应时间、催化剂的配方和制造方法,有时还会遇到搅拌速率和反应器类型等多种因素对优化目标都会有重要影响。这些因素之间,常常相互影响,不能简化为对单因素优选进行考察。将这些因素研究的条件列成表格,把各种可能的搭配逐一进行实验,工作量实在太大,甚至在事实上无法进行。这就需要一种科学的实验设计方法,通过特定安排的一些实验,判断出哪些因素是显著的,哪些是不够显著的,进而抓住主要矛盾,确定最佳工艺条件。正交实验设计,或称正交实验法,就是处理这类问题的有力工具。

正交表的定义:

设 $A$ 是 $n \times k$ 的矩阵,它的第 $j$ 列元素由数字1, 2, 3, ...,  $m$ 所构成(或者方便起见,可用其他符号来代替这些数字),如果矩阵的任意两列都搭配均衡,则称 $A$ 是一个正交表。

由正交表的定义直接推出两个性质:

- 1) 每一列中的各水平出现的次数相同;
- 2) 任意两列所构成的水平对中,每个水平对重复出现次数相同。

### 3.1 正交实验表的设计

根据实验的要求,排出因素(或称因子数),排出



定位级数(或称水平数)。然后,选用相应的正交表。按正交表的安排进行实验。最后,根据实验结果,对诸因素影响的显著程度和顺序作出判断。

正交实验设计的一般步骤为:

- 第一步 明确实验目的,确定考核指标。
- 第二步 选取因素与水平,制定因素水平表。
- 第三步 选择合适的正交表,进行表头设计。
- 第四步 确定实验方案,做实验,填数据。
- 第五步 计算分析实验结果,选取优化方法。
- 第六步 验证实验,最后制定实验方案。

### 3.2 常用正交实验表

二水平正交表:  $L_4(2^3)$ ,  $L_8(2^7)$ ,  $L_{12}(2^{11})$ ,  $L_{16}(2^{15})$ ,  $L_{20}(2^{19})$ ,  $L_{24}(2^{23})$ ,  $L_{28}(2^{27})$ 。

三水平正交表:  $L_9(3^4)$ ,  $L_{27}(3^{13})$ ,  $L_{81}(3^{40})$ ,  $L_{243}(3^{121})$ 。

四水平正交表:  $L_{16}(4^3)$ ,  $L_{64}(4^{21})$ 。

五水平正交表:  $L_{25}(5^4)$ ,  $L_{125}(5^{11})$ 。

七水平正交表:  $L_{49}(7^3)$ 。

常用的混合型正交表:  $L_8(4 \times 2^3)$ ,  $L_{16}(4 \times 2^{12})$ ,  $L_{16}(4^2 \times 2^9)$ ,  $L_{16}(4^3 \times 2^6)$ ,  $L_{27}(9 \times 3^3)$  等。

### 3.3 方差分析和显著因素

方差分析是把实验观测数据分解为各个影响因素的波动和误差波动,然后将它们的平均波动进行比较,中心点就是把观测的数据总的波动分解为反应因素水平变化引起的波动和反应试验误差引起波动两部分。前者是由于因素本身的离散性而存在的方差,简称为产品方差,是产品所固有的;后者是由于实验误差(也称残差)而引起的方差,简称实验方差,是由试验的随机因素引起的。方差分析就是把观测数据的总的偏差平方和( $S_T$ )分解为反映必然性的各个因素的偏差平方和( $S_A, S_B, \dots, S_N$ )与反映偶然性的误差偏差平方和( $S_e$ ),并计算比较它们的平均偏差平方和,以找出对试验观测数据起决定性影响的因素(即显著性或者高度显著性因素)作为进行定量分析判断的依据。

### 3.4 最优实验条件和平均值估计

决定最优试验条件时,必须考虑成本差与减少不良的效应间隙额。在没有给出成本差时,要对不良率小的水平组合估计总体平均值  $\mu$ 。对非显著性的组合采用成本低的水平。

### 3.5 客车轮胎寿命实验

为了做客车轮胎的寿命实验,取轮胎胎号种类  $A$ 、碳粒粒度分布  $B$ 、碳粒数量  $C$ 、添加剂种类  $D$ 、加压温

度  $E$  等五个二水平因素进行轮胎生产。最重要的因素为  $A$ ,其次为  $B$ 、 $C$ 。因而已知存在  $A \times B$ 、 $A \times C$ ,此外四辆客车  $R_1$ 、 $R_2$ 、 $R_3$ 、 $R_4$  可同时进行寿命实验,这些车各自独立进行实验,不做实验上的限制,实验用的轮胎做出配列,关于实验中往何处运行,运行多少公里等都不做实验限制,在通常状态下运行即可。因而,四辆车之间有运行条件的差异。

其次由于轮胎位置会造成实验条件的不同,后面的轮胎比前面的轮胎的条件差,前面的轮胎由于汽车操纵的不对称,司机的位置不处在正中,道路条件等也会产生差异。

因此要考虑这些轮胎的四种位置  $V_1$ (前右)、 $V_2$ (前左)、 $V_3$ (后右)、 $V_4$ (后左)之间的条件差,我们希望的配列是上述五因素的总效应与  $A \times B$ 、 $A \times C$  都没有不利条件的配列。 $V_1$  之间的差是客观存在的。所以  $R_1$ 、 $V_1$  间的差,实际上对  $A$ 、 $B$ 、 $C$ 、 $D$ 、 $E$  的主效应与  $A \times B$ 、 $A \times C$  都未附有不利条件。重要的是如何安排好实验的配列。

实验的目的是求出各因素  $A$ 、 $B$ 、 $C$ 、 $D$ 、 $E$  的信息,不是找出  $R$  与  $V$  的差异。也就是说因素  $A$ 、 $B$ 、 $C$ 、 $D$ 、 $E$  与  $R$ 、 $V$  间存在性质差异。因素  $A$ 、 $B$ 、 $C$ 、 $D$ 、 $E$  为控制因素,确定这些因素为选择最宜水平进行实验。4 辆车  $R_1$ 、 $R_2$ 、 $R_3$ 、 $R_4$  用同一水平进行实验,实验个数为 4 个,以 4 个为界限的因素为区组因素。区组因素虽然也有水平,但其水平没有重要技术意义。对于这样的因素来讲,就连调节效应也没有,其效应是为了不使控制因素间相互影响。车的位置因素  $V$  为标示因素,虽然它具有水平,但无需选用最好的水平。从区组因素与标示因素的说明可以了解,  $R$  与  $V$  的情报是不重要的,是为了区别清楚控制因素的效应而取用的。

表 46.1-2 自由度的分配

要 因	$f$
$A$	1
$B$	1
$C$	1
$D$	1
$E$	1
$A \times B$	1
$A \times C$	1
$R$	3
$V$	3

配列考虑,首先整理取用的因素及其自由度,确定正交表的大小(如表 46.1-2 所示)。因素为二水平,全部自由度为 13,因此宜选  $L_{16}$ (自由度为 15)的正交表,根据因素是二水平为主,还是三水平为主,来决定采取二水平系正交表还是三水平系正交表。 $R$  与  $V$  都是四水平的,对于二水平正交表作四水平列方面,需要二水平的三列。可以利用点线图来表示所要求的主效应与交互作用。和要求的点线图完全相同,标准形点线图-

要是找不到的，所以要从适当的标准点线图通过适当的变形作出相应要求的点线图（如图 46.1-1、46.1-2、46.1-3 所示）。

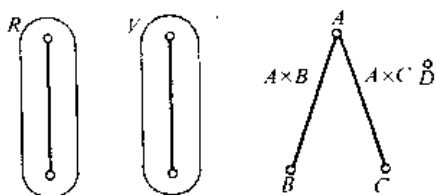


图 46.1-1 要求信息的点线图

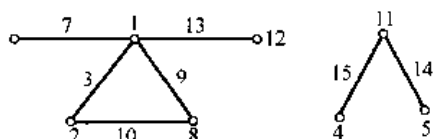


图 46.1-2 点线图 (3)

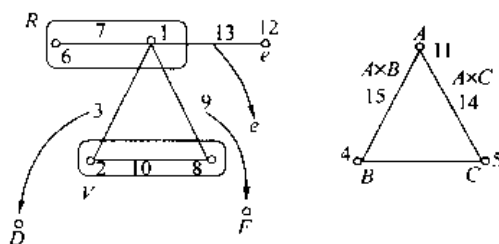


图 46.1-3 利用点线图 (3) 的配列

从以上可知，得到正交表的配列。表 46.1-3 中的 No. 10 实验是利用  $A_1$ 、 $B_1$ 、 $C_2$ 、 $D_2$ 、 $E_1$  方法制造的，装于汽车  $R_2$  的  $V_2$  上的轮胎。但当生产 32 个轮胎时，对于  $A \sim E$  来说，按表中 16 种实验，每一种实验生产 2 个轮胎，可供八辆车来进行实验。八辆车按照下述情况安排：

$R_2$  对  $R_6$                        $R_1$  对  $R_7$

表 46.1 3 实验配列

要因 No.	A	B	C	D	E	A×C	A×B	R (车)	V (位置)	e	e	K <sub>1</sub>	K <sub>2</sub>	K <sub>3</sub>
1	1	1	1	1	1	1	1	$R_1R_3$	$V_1$	1	1	=	=	=
2	2	1	1	1	2	2	2	$R_1R_3$	$V_2$	2	2	=	=	=
3	1	2	2	1	1	2	2	$R_2R_5$	$V_1$	2	2	∴	∴	∴
4	2	2	2	1	2	1	1	$R_2R_5$	$V_2$	1	1	∴	∴	∴
5	2	1	1	2	1	2	2	$R_2R_5$	$V_3$	1	1	∴	∴	∴
6	1	1	1	2	2	1	1	$R_2R_5$	$V_4$	2	2	∴	∴	∴
7	2	2	2	2	1	1	1	$R_1R_3$	$V_3$	2	2	∴	∴	∴
8	1	2	2	2	2	2	2	$R_1R_3$	$V_4$	1	1	∴	∴	∴
9	2	1	2	2	2	1	2	$R_3R_7$	$V_1$	1	2	∴	∴	∴
10	1	1	2	2	1	2	1	$R_3R_7$	$V_2$	2	1	∴	∴	∴
11	2	2	1	2	2	2	1	$R_4R_8$	$V_1$	2	1	∴	∴	∴
12	1	2	1	2	1	1	2	$R_4R_8$	$V_2$	1	2	∴	∴	∴
13	1	1	2	1	2	2	1	$R_4R_8$	$V_3$	1	2	∴	∴	∴
14	2	1	2	1	1	1	2	$R_4R_8$	$V_4$	2	1	∴	∴	∴
15	1	2	1	1	2	1	2	$R_3R_7$	$V_1$	2	1	∴	∴	∴
16	2	2	1	1	1	2	1	$R_3R_7$	$V_4$	1	2	=	=	=
	11	4	5	3	9	14	15	1 6 7	2 8 10	12	13			
	列	列	列	列	列	列	列	列	列	列	列			

$R_2$  对  $R_6$                        $R_1$  对  $R_7$

即可以再增加  $R_5$ 、 $R_4$ 、 $R_7$ 、 $R_8$  四辆车。

今按照表 46.1-3 中的  $I_{16}$  条件，各生产 2 个轮胎，按为  $R_1R_2 \dots R_7R_8$  八辆车做实验。轮胎特性中的动态不平衡量的数据表示见表 46.1-3，是轮胎生产后立即调查的不平衡量；其后把达到平衡的轮胎装在汽车上，令其自由运行。

$K_1$ —汽车装上轮胎前；

$K_2$ —运行了一个月以后；

$K_3$ —运行了两个月以后。

在上述时间把轮胎卸下，测量其不平衡量。数据总

数有  $16 \times 2 \times 3 = 96$  个。

### 3.6 喷淋洗涤实验

#### 3.6.1 配列与数据

求取喷淋洗涤最宜条件，第一次实验中把 48 个清洗对象分为两组，24 个一组。

$R_1$ —污物较少的一组；

$R_2$ —污物较多的一组。

并且控制因素取：

A: 洗剂 [牌号]                       $A_1=N$ ,  $A_2=C$ ,

$A_3=K, A_4=S$   
 $B$ : 浓度 [%]  $B_1=2, B_2=4$   
 $C$ : 温度 [C]  $C_1=50, C_2=80$   
 $D$ : 压力 [kg/mm<sup>2</sup>]  $D_1=3, D_2=5$

把上述四因素配列于  $A, B, C, D$ , 按 8 种条件从  $R_1, R_2$  中各取三个, 对此 6 个做了连续的洗涤实验。清洗 45s 后在外观上分上、中、下三组, 结果见表 46.1-4。

表 46.1-4 配列与数据

因子列 No.	A			B 4	C 5	D 6	e 7	R <sub>1</sub>			R <sub>2</sub>		
	1	2	3					上	中	下	上	中	下
1	1	1	1	1	1	1	0	3	0	0	1	2	
2	1	2	2	2	2	2	3	0	0	1	2	0	
3	2	1	1	2	2	1	1	2	0	0	1	2	
4	2	2	2	1	1	3	0	0	2	1	0	2	
5	3	1	2	1	2	2	1	0	0	0	0	3	
6	3	2	1	2	1	2	1	0	0	2	1	1	
7	4	1	2	2	1	1	2	0	0	2	1	0	
8	4	2	1	1	2	2	1	0	0	3	0	1	

3.6.2 方差分析

数据是技术分类值, 分为上、中、下三组, 根据累加法做出分析。当进入分析前, 先求出方差分析表的因素与自由度的分配, 见表 46.1-5。

表 46.1-5 自由度的分配

要 因	f
A	3×2=6
B	1×2=2
C	1×2=2
D	1×2=2
e <sub>1</sub>	1×2=2
R	1×2=2
A×R	3×2=6
B×R	1×2=2
C×R	1×2=2
D×R	1×2=2
e <sub>2</sub>	1×2=2
e <sub>3</sub>	32×2=64
计	47×2=94

表 46.1-6 变换与累积频数的数据

因子 No.	A	B	C	D	e	R <sub>1</sub>			R <sub>2</sub>		
						I	II		I	II	
1	1	1	1	1	1	0	3	0	1	2	1
2	1	2	2	2	2	3	3	1	2	0	3
3	2	1	1	2	2	1	3	0	1	2	1
4	2	2	2	1	1	3	3	2	3	0	3
5	3	1	2	1	2	2	3	0	0	0	3
6	3	2	1	2	1	2	3	0	2	1	2
7	4	1	2	2	1	1	3	0	2	1	2
8	4	2	1	1	2	2	3	0	3	0	3

如  $R$  的情况, 于正交表外部的因素为外部因素, 不仅求其主效应, 一般还要求出与正交分配表各因素的交互作用。顺次计算表 46.1-6 中各因素的波动。首先把数据变换为累积频数, 做出辅助表 46.1-7。

表 46.1-7 辅助表

	R <sub>1</sub>		R <sub>2</sub>		小计	
	I	II	I	II	I	II
A <sub>1</sub>	3	6	1	4	4	10
A <sub>2</sub>	4	6	2	4	6	10
A <sub>3</sub>	4	6	0	2	4	8
A <sub>4</sub>	3	6	0	5	3	11
B <sub>1</sub>	4	12	0	4	4	16
B <sub>2</sub>	10	12	3	11	13	23
C <sub>1</sub>	5	12	0	7	5	19
C <sub>2</sub>	9	12	3	8	12	20
D <sub>1</sub>	7	12	2	7	9	19
D <sub>2</sub>	7	12	1	8	8	20
e <sub>1</sub>	6	12	2	8	8	20
e <sub>2</sub>	8	12	1	7	9	19
计	14	24	3	15	17	39

$$\bar{W}_1 = \frac{1}{p_1(1-p_1)} = \frac{1}{\frac{17}{48} \cdot (1 - \frac{17}{48})} = \frac{48^2}{17(48-17)} = 4.37$$

$$\bar{W}_2 = \frac{48^2}{39(48-39)} = 6.56$$

$$S_A = \frac{(4^2+6^2+4^2+3^2) \times \bar{W}_1 + (10^2+10^2+8^2+11^2) \times \bar{W}_2}{12} - \frac{17^2 \times \bar{W}_1 + 39^2 \times \bar{W}_2}{48}$$

$$= 238.51 - 234.18 = 4.33$$

$$S_B = \frac{(4-13)^2 \times 4.37 + (16-23)^2 \times 6.56}{48} = 14.07$$

$$S_C = \frac{(5-12)^2 \times 4.37 + (19-20)^2 \times 6.56}{48} = 4.60$$

$$S_D = \frac{(7-8)^2 \times 4.37 + (19-20)^2 \times 6.56}{48} = 0.23$$

$$S_{e_1} = \frac{(8-9)^2 \times 4.37 + (20-19)^2 \times 6.56}{48} = 0.23$$

$$S_R = \frac{(14-3)^2 \times 4.37 + (24-15)^2 \times 6.56}{48} = 22.09$$

$$S_{A \times R} = \frac{(3^2+4^2+\dots+0^2) \times 4.37 + (6^2+6^2+\dots+5^2) \times 6.56}{6}$$

$$- \frac{17^2 \times 4.37 + 39^2 \times 6.56}{48} - S_A - S_R$$

$$= -3.59$$

$$S_{B \times R} = \frac{(4+3-10-0)^2 \times 4.37 + (12+11-12-4)^2 \times 6.56}{48}$$

$$= 7.52$$

$$S_{C \times R} = \frac{(5+3-0-9)^2 \times 4.37 + (12+8-7-12)^2 \times 6.56}{48}$$

$$= 0.23$$

$$S_{D \times R} = \frac{(7+1-2-7)^2 \times 4.37 + (12+8-7-12)^2 \times 6.56}{48}$$

$$= 0.23$$

$$S_{e_2} = \frac{(6+1-2-8)^2 \times 4.37 + (12+7-8-12)^2 \times 6.56}{48}$$

-0.96

$S_7 = 48 \times 2 = 96$

$S_{\epsilon_3} = S_T - S_A - S_B - \dots - S_{\epsilon_2}$

$= 96 - 1.33 - 14.07 - 4.60 - 0.23 - 0.23$   
 $- 22.09 - 3.59 - 7.52 - 0.23 - 0.96 - 0.23$   
 $= 37.92$

整理分析后的方差分析表,见表 46.1-8。

表 46.1-8 方差分析表

要因	f	S	V	$F_0$	$\rho[\%]$
A	6	4.33	0.72(○)		
B	2	14.07	7.04	12.68	13.50
C	2	4.60	2.30	4.14	3.64
D	2	0.23	0.12(○)		
$\epsilon_1$	2	0.23	0.12(○)		
R	2	22.09	11.04	19.89	21.85
A×R	6	3.59	0.60(○)		
B×R	2	7.52	3.76	6.77	6.68
C×R	2	0.23	0.12(○)		
D×R	2	0.23	0.12(○)		
$\epsilon_2$	2	0.96	0.48(○)		
$\epsilon_3$	64	37.92	0.592(○)		
计 (○记号合计)	94 (86)	96.00 (47.72)			100 (54.33)

注:○记号意义为较小因素,可忽略。

3.6.3 显著因素估计

由方差分析结果得出,显著因素为 B、C、R、B×R, 这些因素效应的估计归纳见表 46.1-9。

表 46.1-9 各因素效应的估计

	上	中	下	计
$B_1$	16.7	50.0	33.3	100.0
$B_2$	54.2	41.6	4.2	100.0
平均	35.4	45.8	18.8	100.0
	上	中	下	计
$C_1$	20.8	58.4	20.8	100.0
$C_2$	50.0	33.3	16.7	100.0
平均	35.4	45.8	18.8	100.0
	上	中	下	计
$R_1$	58.3	41.7	0.0	100.0
$R_2$	12.5	50.0	37.5	100.0
平均	35.4	45.8	18.8	100.0
	上	中	下	计
$B_1R_1$	33.3	66.7	0.0	100.0
$B_1R_2$	0.0	33.3	66.7	100.0
$B_2R_1$	83.3	16.7	0.0	100.0
$B_2R_2$	23.0	66.7	8.3	100.0
平均	35.4	45.8	18.8	100.0

3.6.4 最宜条件的决定与预测

对于这一问题,R 不是控制因素,是作为标示因素来考虑的,所以必须作出每一  $R_1R_2$  平均工序的估计。

其最宜条件既是  $R_1R_2$  又是  $B_2C_2$ 。考虑 B×R 为显著因素进行下述估计。

(1)  $R_1$  的情况

I 组

$\hat{\mu}_1$  的分贝值 =  $\overline{B_1R_1}$  的分贝值 - T 的分贝值  
 $= 83.3\%$  的分贝值 + 50.0%  
 的分贝值 - (35.4% 的分贝值)  
 $= 6.979 + 0.000 - (-2.611)$   
 $= 9.590 \text{ (dB)}$

$\hat{\mu}_1 = 90.1\%$

II 组

$\hat{\mu}_1$  的分贝值 = 100.0% 的分贝值 + 83.3% 的分贝值 - (81.2% 的分贝值)  
 $= \infty$

$\mu_1 = 100.0\%$

于是可以预想  $R_1$  的情况,上为 90.1%,中为 9.9%,下为 0%。

(2)  $R_2$  的情况

I 组

$\hat{\mu}_1$  的分贝值 =  $\overline{B_2R_2}$  的分贝值 +  $C_2$  的分贝值  
 $= 25.0\%$  的分贝值 + 50.0% 的分贝值 - (35.4% 的分贝值)  
 $= [-4.770 - 0.000 - (-2.611)] \text{ (dB)}$   
 $= -2.159 \text{ (dB)}$

$\hat{\mu}_1 = 37.8 \pm \sqrt{3.96 \times 0.555 \times 0.378 \times 0.622 \times \frac{10}{96}}$   
 $\times 100$   
 $= (37.8 \pm 23.2)\%$

II 组

$\hat{\mu}_1$  的分贝值 = 91.7% 的分贝值 + 83.3% 的分贝值 - (81.2% 的分贝值)  
 $= 10.433 + 6.979 - (-0.730)$   
 $= 11.058 \text{ (dB)}$

$\hat{\mu}_1 = 92.7 \pm \sqrt{3.96 \times 0.555 \times 0.927 \times 0.073 \times \frac{10}{96}}$   
 $\times 100$   
 $= (92.7 \pm 3.9)\%$

于是可以预想  $R_2$  的情况,上为 37.8%,中为 54.9%,下为 7.3%。

3.7 铸铝机部件的实验

3.7.1 配列与数据

为了防止铸铝零件产生气孔,取以下因素:

A: 材料

A: 现行材料,  $A_2 = N$ ,

$A_1 = S$ ,  $A_4 = K$ ,

$A_2 = SV$ ,  $A_5 = 0$ ,

$A_7 = DR$

B: 金属模

$B_1 =$ 新的,  $B_2 =$ 旧的

C: 砂型涂料工艺

$C_1 =$ 现行工艺,  $C_2 =$ 新工艺

D: 铸铝温度

$D_1 =$ 现行,  $D_2 =$ 比现行增加 30°C

E: 脱气时间

$E_1 =$ 现行,  $E_2 =$ 现行的 2 倍

F: 镇静时间

$F_1 =$ 现行,  $F_2 =$ 现行的 3 倍

G: 砂型涂料材料 (泥芯)

$G_1 =$ 现行,  $G_2 =$ 新的

将这些因素按正交表  $L_{16}$  进行配列, 因素 A 再取拟水平  $A_1$  作为第八水平。对每次实验中所铸得的产品按气孔的多少进行分类, 将每类的产品个数作为实验数据。(见表 46.1-10 右侧)

表 46.1-10 配列与数据

因素列 No.	A	D	B	G	e	C	e	E	F	多	中	少	无	总计
	1234567	8	9	10	11	12	13	14	15					
1	1	1	1	1	1	1	1	1	1	20	27	49	141	237
2	1	2	2	2	2	2	2	2	2	5	8	15	117	145
3	2	1	1	1	1	2	2	2	2	0	1	7	109	117
4	2	2	2	2	2	1	1	1	1	10	31	67	121	229
5	3	1	1	2	2	1	1	2	2	5	14	34	127	180
6	3	2	2	1	1	2	2	1	1	1	1	5	119	126
7	1	1	1	2	2	2	2	1	1	0	0	5	138	143
8	4	2	2	1	1	1	1	2	2	1	3	12	194	210
9	1'	1	2	1	2	1	2	1	2	29	64	110	138	341
10	1'	2	1	2	1	2	1	2	1	3	6	11	130	150
11	5	1	2	1	2	2	2	1	2	0	1	4	150	155
12	5	2	1	2	1	1	2	1	2	0	4	7	152	163
13	6	1	2	2	1	1	2	2	1	9	8	92	122	231
14	6	2	1	1	2	2	1	1	2	0	0	2	129	131
15	7	1	2	2	1	2	1	1	2	0	0	3	138	141
16	7	2	1	1	2	1	2	2	1	3	10	33	186	232

3.7.2 方差分析

由表 46.1-10 可知, 数据分为四组, 是计数分类值数据。依据分类项目, 采用累积法进行数据分析, 首先列出表 46.1-11 左侧的累积频数表。

表 46.1-11 累积频数与出现率数据

	累积频数				出现率数据			
	I	II	III	IV	I	II	III	IV
1	20	47	96	237	0.084	0.198	0.405	1.000
2	5	13	28	145	0.034	0.090	0.193	1.000
3	0	1	8	117	0.000	0.009	0.068	1.000
4	10	41	108	229	0.044	0.179	0.472	1.000
5	5	19	53	180	0.028	0.106	0.294	1.000
6	1	2	7	126	0.008	0.016	0.056	1.000
7	0	0	5	143	0.000	0.000	0.035	1.000
8	1	4	16	210	0.005	0.019	0.076	1.000
9	29	93	203	341	0.085	0.273	0.595	1.000
10	3	9	20	150	0.020	0.060	0.133	1.000
11	0	1	5	155	0.000	0.006	0.032	1.000
12	0	4	11	163	0.000	0.024	0.067	1.000
13	9	17	109	231	0.039	0.074	0.472	1.000
14	0	0	2	131	0.000	0.000	0.015	1.000
15	0	0	3	141	0.000	0.000	0.021	1.000
16	3	13	46	232	0.013	0.056	0.198	1.000
总计	86	264	720	2931	0.360	1.110	3.132	16.000

然而每次实验的生产的的产品数, 即重复数是不等。对这样的数据, 把累积频数变成出现铝后用累积法进行分析。分析的数据分为三组, 如表 46.1-11 右边所示

接着列出辅助表 (见表 46.1-12)。

表 46.1-12 辅助表

	I	II	III
$A_1$	0.223	0.621	1.326
$A_2$	0.044	0.188	0.540
$A_3$	0.036	0.122	0.350
$A_4$	0.005	0.019	0.111
$A_5$	0.000	0.030	0.099
$A_6$	0.039	0.074	0.487
$A_7$	0.013	0.056	0.219
$B_1$	0.145	0.453	1.215
$B_2$	0.215	0.657	1.917
C	0.298	0.929	2.579
$C_2$	0.062	0.181	0.553
$D_1$	0.236	0.666	1.922
$D_2$	0.124	0.444	1.210
$E_1$	0.221	0.690	1.666
$E_2$	0.139	0.420	1.466
$F_1$	0.208	0.589	1.803
$F_2$	0.152	0.521	1.329
$G_1$	0.195	0.577	1.445
$G_2$	0.165	0.538	1.687
$(11)_1$	0.156	0.400	1.298
$(11)_2$	0.204	0.710	1.834
$(13)_1$	0.181	0.568	1.449
$(13)_2$	0.179	0.542	1.684
总计	0.360	1.110	3.132

权数计算如下:

$$\left. \begin{aligned} W_1 &= \frac{16^2}{0.36 \times 15.640} = \frac{256}{5.630} = 45.47 \\ W_2 &= \frac{16^2}{1.110 \times 14.890} = \frac{256}{16.528} = 15.49 \\ W_3 &= \frac{16^2}{3.132 \times 12.868} = \frac{16^2}{40.303} = 6.35 \end{aligned} \right\}$$

然后利用权数求各个因素的波动。

$$CF = \frac{0.360^2 \times W_1 + 1.110^2 \times W_2 + 3.132^2 \times W_3}{16} = 5.454$$

$$S_A = \frac{1}{4} (0.223^2 \times W_1 + 0.621^2 \times W_2 + 1.326^2 \times W_3) + \frac{1}{2} [(0.044^2 + 0.036^2 + \dots + 0.013^2) \times W_1 + (0.188^2 + \dots + 0.056^2) \times W_2 + (0.540^2 + \dots + 0.219^2) \times W_3] - CF = 4.850 + 2.868 - 5.454 = 2.264 \quad (f=18)$$

$$S_B = \frac{1}{16} [(0.145 - 0.215)^2 \times W_1 + (0.453 - 0.657)^2 \times W_2 + (1.215 - 1.917)^2 \times W_3] = \frac{3.997}{16} = 0.250 \quad (f=3)$$

$$S_C = \frac{1}{16} [(0.298 - 0.062)^2 \times W_1 + (0.929 - 0.181)^2 \times W_2 + (2.579 - 0.553)^2 \times W_3] = \frac{37.264}{16} = 2.329 \quad (f=3)$$

$$S_D = \frac{1}{16} [(0.236 - 0.124)^2 \times W_1 + (0.666 - 0.444)^2 \times W_2 + (1.922 - 1.210)^2 \times W_3] = \frac{4.553}{16} = 0.285 \quad (f=3)$$

$$S_E = \frac{1}{16} [(0.221 - 0.139)^2 \times W_1 + (0.690 - 0.420)^2 \times W_2 + (1.666 - 1.466)^2 \times W_3] = \frac{1.689}{16} = 0.106 \quad (f=3)$$

$$S_F = \frac{1}{16} [(0.208 - 0.152)^2 \times W_1 + (0.589 - 0.521)^2 \times W_2 + (1.803 - 1.329)^2 \times W_3] = \frac{1.641}{16} = 0.103 \quad (f=3)$$

$$S_G = \frac{1}{16} [(0.195 - 0.165)^2 \times W_1 + (0.577 - 0.533)^2 \times W_2 + (1.445 - 1.687)^2 \times W_3] = \frac{0.443}{16} = 0.028 \quad (f=3)$$

$$S_{T_1} = (0.084^2 + 0.034^2 + \dots + 0.013^2) \times W_1 +$$

$$(0.0918^2 + 0.090^2 + \dots + 0.056^2) \times W_2 + (0.405^2 - 0.193^2 + \dots + 0.198^2) \times W_3 - CF = 11.090 - 5.454 = 5.636 \quad (f=45)$$

$$S_{e_1} = S_{T_1} - S_A - S_B - S_C - S_D - S_E - S_F - S_G = 0.271 \quad (f=9)$$

下面求二次误差波动  $S_{e_2}$ 。设  $S'_{e_2}$  是波动第 1, 2, …, 16 次重复实验之间的波动, 则  $S'_{e_2}$  可以按照各组的波动加权以后相加得到, 即:

$$S'_{e_2} = (\text{I 组全部测定值的平方和} - \text{实验间的波动}) \times W_1 + (\text{II 组全部测定值的平方和} - \text{实验间的波动}) \times W_2 + (\text{III 组全部测定值的平方和} - \text{实验间的波动}) \times W_3 = \left[ 86 - \left( \frac{20^2}{237} + \frac{5^2}{145} + \dots + \frac{3^2}{232} \right) \right] \times W_1 + \left[ 264 - \left( \frac{47^2}{237} + \frac{13^2}{145} + \dots + \frac{13^2}{232} \right) \right] \times W_2 + \left[ 720 - \left( \frac{96^2}{237} + \frac{28^2}{145} + \dots + \frac{46^2}{232} \right) \right] \times W_3 = 9691.582 \quad (f=8745)$$

然而将现在所求得  $S'_{e_2}$  除以重复数的调和平均数, 是因为其他各因素的波动  $S_A, S_B, \dots$  都是除以重复数后的平均值的波动。调和平均值  $\bar{r}$  计算如下:

$$\frac{1}{\bar{r}} = \frac{1}{16} \left( \frac{1}{237} + \frac{1}{145} + \frac{1}{117} + \dots + \frac{1}{232} \right)$$

于是

$$\bar{r} = 168.73$$

由此可知

$$S_{e_2} = \frac{S'_{e_2}}{\bar{r}} = \frac{9691.582}{168.73} = 57.438$$

总结以上的计算结果得方差分析表, 见表 46.1-13。

表 46.1-13 方差分析表

要 因	f	S	V
A	18	2.264	0.126
B	3	0.250	0.083○
C	3	2.329	0.776
D	3	0.285	0.095○
E	3	0.106	0.035○
F	3	0.103	0.034○
G	3	0.028	0.009○
$e_1$	9	0.271	0.030
带有○的 ( $e_1$ )	(24)	(1.043)	(0.044)
$e_2$	8745	57.438	0.0066
总计	8790	63.074	

在方差分析表中, 用  $e_2$  来检验,  $e_1$  是显著的。这时把效应小的因素并入中  $e_1$  去, 组成新的误差方差, 用以检验各因素的效应。注意的是, 不要把  $S_{e_2}$  归并调。

$$V_{c_1} = \frac{S_{c_1} + S_B + S_D + S_A - S_k - S_c}{9 + 3 + 3 + 3 + 3 - 3}$$

$$= \frac{1.043}{24} = 0.044$$

这时 A、C 是显著因素。

估计与最佳条件

对因素 A、C 进行估计，I 组的  $\bar{A}_1$  估计如下：

$$\bar{A}_1 = \frac{0.223}{4} \pm \sqrt{F_{24} (0.005) \times \frac{V_{c_1}}{n_r}}$$

式中  $F_{24} (0.05)$  是分子自由度为 1，分母自由度（归并后的误差方差自由度）为 24 的 F 分布在 5% 的值，它等于 4.26，由于方差分析表中的  $V_c$  是原来误差方差的  $\frac{1}{r}$  倍，所以必须对  $V_c$  作扩大  $r$  倍的修正。而且在累积法中乘上了作为权数的  $p(1-p)$  的倒数，所以方差分析表中的误差方差是与  $p(1-p)$  成比例的系数，即所谓比例常数。因此估计值是  $p$ ，用  $\bar{p}(1-\bar{p})$  乘以方差分析表中的误差方差项，去估计  $\bar{p}$  领域的误差方差。也可以用下式作为 I 组共同的误差方差：

$$V_{c_1} = V_c \times p \times (1-p) \times r = \frac{V_c \times r}{W_1}$$

设所有的实验都有调和平均值  $\bar{r}$  的重复，则  $A_1$  时的  $n_r = 4\bar{r}$ 。

于是

$$A_1 = \frac{0.223}{1} + \sqrt{F_{24} (0.05) \times \frac{V_{c_1}}{n_r}}$$

$$= 0.056 - \sqrt{4.26 \times \frac{V_{c_1} \times r}{W_1} \times \frac{1}{4r}}$$

$$= 0.056 - \sqrt{4.26 \times \frac{0.044}{45.47} \times \frac{1}{4}}$$

$$= 0.056 - 0.032$$

$$\bar{A}_1 = \frac{0.044}{2} \pm \sqrt{4.26 \times \frac{0.044}{45.47} \times \frac{1}{2}}$$

$$\therefore 0.022 \pm 0.045$$

同样计算得到

$$\left. \begin{aligned} \bar{A}_2 &= 0.018 \pm 0.045 \\ \bar{A}_4 &= 0.002 \pm 0.045 \\ \bar{A}_5 &= 0.000 \pm 0.045 \\ \bar{A}_6 &= 0.020 \pm 0.045 \\ \bar{A}_7 &= 0.006 \pm 0.045 \end{aligned} \right\}$$

II 组，III 组也可以同样地计算得出。

## 第 2 章 基本实验设计方法

### 1 裂区法设计

所谓分割法是分阶段地使不同的因素具有不同的实际实验次数的一种方法，不是对各因素同样进行实验随机化。

#### 1.1 单元裂区设计和因素分类

在同一实验中的因素，通常按照它们水平的难易程度分成如下三组：

- 1) 一次因素：水平较难改变的因素。
- 2) 二次因素：除一次三次因素外的因素。
- 3) 三次因素：水平容易改变的因素。

因素分组同样可以分为两组，四组。在进行实验设计之前，必须对各个因素的水平改变，实际需要的时间，经济上损失如何等作出充分的研究，并且将它们按上述方法进行分组。

#### 1.2 正交表的列的分组

正交表最后一行表示表中各列属于哪一群。以  $L_{16}(2^{15})$  为例，各列的分群如表 46.2-1 所示。即第 1 列为第 1 群，第 2, 3 列为第 2 群，第 4, 5, 6, 7 列为第 3 群，第 8, 9, 10, 11, 12, 13, 14, 15 列是第 4 群。第 2 群与第 1 群相比是高次。这样分群，在各因素的水平

更换没有难易差别的情况下是没有意义的，因为计算只在主效应与交互作用之间时，不必进行列的选择，只按随机顺序进行即可，而 15 个列的排列顺序就没有特别的意义，因此分群也没有意义了。现在无论哪个水平的都要同样做 8 次反复实验，所以如果有些因素更换水平比较困难，这样做实验就要花费许多不必要的时间和经费。这样把正交表的各列分成若干群，为了每个群与因素的每个组相对应，使一次、二次、三次因素的实验反复次数不同，实验就相对容易些。

下述是分割法中的交互作用规则：

交互作用的分割规则 a：属于不同群的两列的积（交互作用）一定组成高次群中的列。

交互作用的分割规则 b：属于同一群的两列的积（交互作用），其全部或部分一定属于次数低于它的群。属于二水平的情况，则全部属于低于它的群，属于  $p$  水平的情况，对应于  $p-1$  的各列中，只有一个列属于次数低于它的群。

现在用正交表  $L_{16}(2^{15})$  来说明以上两种规则。第 3 列在第 2 群，第 9 列在第 4 群，所以按两列之间的交互作用表可知，上述两列的积构成第 10 列，与第 9 列一样属于第 4 群。而同为第 4 群的第 8 列和第 9 列，它们的积属于低次的第 2 群的第 2 列。

表 46.2-1 正交表  $L_{16}(2^{15})$

列 No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1
群	1 群		2 群		3 群			4 群							



1.3 薄板实验

1.3.1 配列与数据

在层压板的实验配列中，一次因素 B 安排到第 1~3 列，二次因素 C、D、E 安排到第 4~7 列，三次因素安排到第 8~15 列。

但是由于四水平因素 A 不能全部安排在第 4 群，所以不得不把其中的一个自由度安排到第 7 列。在实

验中低次因素安排到高次群是不可能的，但是相反的安排却是可以的。但这样做精度会降低，所以不理想。因素配列见表 46.2-2。每次实验都生成一块层压板，每块板上各取两个试样进行磨损实验，测得磨损深度。磨损深度减去虚拟平均值 30μm 后，就是表 46.2-2 中的数据。

对于各因素，计算不同水平的合计值，列出辅助表（见表 46.2-3）。

表 46.2-2 因素的配列和数据

因素列 No.	B 1	e <sub>1</sub> 2	e <sub>2</sub> 3	C 4	D 5	E 6	(A) 7	A 8	(A) 8	F 9	G 10	B×G 11	H 12	V 13	e <sub>3</sub> 14	(A) 15	数据 (e <sub>4</sub> )	总计	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	22	12
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	-1	-3	-1
3	1	1	1	2	2	2	2	3	1	1	1	1	2	2	2	2	2	1	3
4	1	1	1	2	2	2	2	4	2	2	2	2	1	1	1	1	-1	-5	-6
5	1	2	2	1	1	2	2	3	1	1	2	2	1	1	2	2	3	0	3
6	1	2	2	1	1	2	2	4	2	2	1	1	2	2	1	1	2	3	5
7	1	2	2	2	2	1	1	1	1	1	2	2	2	2	1	1	25	23	48
8	1	2	2	2	2	1	1	2	2	2	1	1	1	1	2	2	5	1	6
9	2	1	2	1	2	1	2	3	1	2	1	2	1	2	1	2	-14	-14	-28
10	2	1	2	1	2	1	2	4	2	1	2	1	2	1	2	1	-17	-19	-36
11	2	1	2	2	1	2	1	1	1	2	1	2	2	1	2	1	23	24	47
12	2	1	2	2	1	2	1	2	2	1	2	1	1	2	1	?	-3	-1	-4
13	2	2	1	1	2	2	1	1	1	2	2	1	1	2	2	1	20	16	36
14	2	2	1	1	2	2	1	2	2	1	1	2	2	1	1	2	8	7	15
15	2	2	1	2	1	1	2	3	1	2	2	1	2	1	1	2	-21	-20	-41
16	2	2	1	2	1	1	2	4	1	1	1	2	1	2	2	1	-11	-11	-22
总计																		64	

表 46.2-3 辅助表

	B	(2)	(3)	C	D	E	F	G	B×G
第一水平	97	14	23	33	26	-35	49	68	11
第二水平	-33	50	41	31	38	99	15	-4	53
总计	64	64	64	64	64	64	64	64	64
	H		V		(14)		A <sub>1</sub>		173
第一水平	27		30		31		A <sub>2</sub>		13
第二水平	37		34		33		A <sub>3</sub>		63
总计	64		64		64		A <sub>4</sub>		-59

1.3.2 方差分析

(1) 计算修正项

总波动 S<sub>T</sub>，实验间的波动 S<sub>T1</sub>。

$$CF = \frac{64^2}{32} = 128 \quad (f=1)$$

$$S_T = 20^2 + 22^2 + \dots + (-11)^2 - CF = 6008 \quad (f=31)$$

$$S_{T_3} = \frac{1}{2} [42^2 + (-4)^2 + \dots + (-22)^2] - CF = 5967 \quad (f=15)$$

(2) 计算与一次因素有关的波动

$$S_B = \frac{(97+33)^2}{32} = 528 \quad (f=1)$$

$$S_{e_1} = S_{201} + S_{331} = \frac{(14-50)^2 + (23-41)^2}{32}$$

$$= 51 \quad (f=2)$$

一次误差的波动 S<sub>e</sub> 可以从一次单位之间的总波动 S<sub>T1</sub> 减去 S<sub>B</sub>。可以进行验算。

$$S_{T_1} = \frac{1}{8} [(42-1-3-6)^2 + (3+5+48+6)^2 + \dots + (36+15-41-22)^2] - CF = 579 \quad (f=3)$$

(3) 计算与二次因素有关的波动

$$S_C = \frac{(32-31)^2}{32} = 0 \quad (f=1)$$

$$S_D = \frac{(26-38)^2}{32} = 4 \quad (f=1)$$

$$S_F = \frac{(-35-99)^2}{32} = 561 \quad (f=1)$$

此外,四水平因素 A 的效应部分也在二次因素里,设它为  $S_{A'}$ ,则  $A'$  是  $A_1, A_2$  与  $A_3, A_4$  之差的因素

$$S_{A'} = \frac{(186+122)^2}{32} = 2964 \quad (f=1)$$

(4) 计算三次因素有关的波动

因素 A 的效应中的  $A_1, A_2$  之差,  $A_3, A_4$  之差是三次因素。

$$S_{A_1A_2} = \frac{(173-13)^2}{16} = 1600 \quad (f=1)$$

$$S_{A_3A_4} = \frac{(-63+59)^2}{16} = 1 \quad (f=1)$$

$$S_F = \frac{(49-15)^2}{32} = 36 \quad (f=1)$$

$$S_G = \frac{(68+4)^2}{32} = 162 \quad (f=1)$$

$$S_{B \times G} = \frac{(11-53)^2}{32} = 55 \quad (f=1)$$

$$S_H = \frac{(27-37)^2}{32} = 3 \quad (f=1)$$

$$S_V = \frac{(30-34)^2}{32} = 1 \text{ (取整)} \quad (f=1)$$

$$S_{e_1} = \frac{(31-33)^2}{32} = 0 \text{ (取整)} \quad (f=1)$$

这些因素的总计必须与实验间的波动一致。

(5) 四次误差波动  $S_{e_4}$

四次误差波动  $S_{e_4}$ , 即重复实验间误差波动

$$S_{e_4} = \frac{1}{2} [(20-22)^2 + (-1+3)^2 + \dots + (-11+11)^2] = 41 \quad (f=16)$$

因此得方差分析表 (见表 46.2.4)。

表 46.2.4 方差分析表

因素	f	S	V	F <sub>0</sub>
一次因素 B	1	528	528	20.7
一次误差 e <sub>1</sub>	2	51	25.5	11.1
二次因素 { A' C D E	1 1 1 1	2964 0 4 561	2964 0 4 561	1288.7  243.9
二次误差 e <sub>2</sub>	0	0	(-)	
三次因素 { A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> A <sub>4</sub> B G H B×C V	1 1 1 1 1 1 1	1600 1 36 162 3 55 1	1600 1 36 162 3 55 1	695.7 15.6 70.4 23.9
三次误差 e <sub>3</sub>	1	0	0	
四次误差 e <sub>4</sub> (e <sub>2</sub> )	16 (22)	41 (50)	2.6 (2.3)	
总计	31	6008		

由方差分析表可知,用重复误差  $e_1$  去检验  $A_3, A_4, H, V, e_3$  都不显著,所以三次误差方差的存在与否不能肯定。因为 C、D 不是显著的,所以二次误差方差  $e_2$  的存在与否也不能肯定。用由 C、D、 $A_3, A_4, V, H, e_3, e_1$  归并而成的误差检验,一次误差  $e_1$  是显著的。

$$F = \frac{25.5}{2.3} = 11.1$$

因此一次因素 B 就用  $e_1$  来检验:

$$F_{B, e_1} = \frac{V_B}{V_{e_1}} = \frac{528}{25.5} = 20.7$$

这个值用 5% 水平是显著的,但是这时  $e_1$  的自由度只有 2,所以即使 B 不显著但 F 大于 5 时,最好不要忽略 B 的效应。

由此整理后得方差分析表 (见表 46.2-5)。

表 46.2-5 整理后的方差分析表

因素	f	S	V	F <sub>0</sub>	S'	ρ (%)
B	1	528	528.0	20.7	502.5	8.36
e <sub>1</sub>	2	51	25.5	11.1	69.6	1.17
A	3	4565	1521.7	661.6	4558.1	75.87
E	1	561	561.0	243.9	558.7	9.29
F	1	36	36.0	15.6	33.7	0.56
G	1	162	162.0	70.4	159.7	2.66
B×G	1	55	55.0	23.9	52.7	0.88
e <sub>2</sub>	(21)	(49)	(2.3)		72.0	1.21
总计	31	6008			6008.0	100.00

各个纯波动的计算:

$$S'_B = 528 - 25.5 = 502.5$$

$$S'_{e_1} = 51 + 25.5 - 3 \times 2.3 = 69.6$$

$$S'_A = 4565 - 3 \times 2.3 = 4558.1$$

$$S'_E = 561 - 2.3 = 558.7$$

⋮

$$S'_{e_2} = 49 + 10 \times 2.3 = 72.0$$

显然  $A_3, A_4$  的差是显著的,但是因为二次误差,三次误差可以归并到重复误差中,所以 A 就汇总了一个四水平的因素。因此  $A_3, A_4$  之间的差不是显著的,应当归并到  $S_A$  中,并且从归并成的误差  $e_2$  中扣除。

### 1.3.3 因素的估计与最宜条件

估计显著因素的效应。因为  $B \times G$  是显著的,所以必须对 B 与 G 的组合进行估计。由于一次因素 B 的主效应是显著的,所以首先进行估计, B 的主效应加上虚拟平均值 30 后得

$$\left. \begin{aligned} B_1 &= 30 + \frac{97}{16} = 36.1 \pm 5.4 \\ B_2 &= 30 + \frac{-33}{16} = 27.9 \pm 5.4 \end{aligned} \right\}$$

它们的置信界限区间为:

$$\pm\sqrt{F \times V_{e_1} \times \frac{1}{n_e}} = \pm\sqrt{18.51 \times 25.5 \times \frac{1}{16}}$$

$$= \pm 5.4$$

因为一次误差是显著的, 并且  $B$  必须用一次误差去检验, 所以  $B$  的置信界限中使用了二次误差方差  $V_{e_1}$ ,  $F$  的分母自由度也使用  $V_{e_1}$  的自由度, 所以上式中的  $F$  值是分子自由度为 1, 分母自由度为 2 的  $F$  分布在 5% 点的值, 为 18.51。

下面估计二次因素

$$\bar{A}_1 = 30 + \frac{173}{8} = 51.6 \pm 1.1$$

$$\bar{A}_2 = 30 + \frac{13}{8} = 31.6 \pm 1.1$$

$$\bar{A}_3 = 30 + \frac{-63}{8} = 22.1 \pm 1.1$$

$$\bar{A}_4 = 30 + \frac{-59}{8} = 22.6 \pm 1.1$$

它们的置信界限为

$$\pm\sqrt{4.32 \times 2.3 \times \frac{1}{8}} = \pm 1.1$$

$$E_1 = 30 + \frac{-35}{16} = 27.8 \pm 0.8$$

$$E_2 = 30 + \frac{99}{16} = 36.2 \pm 0.8$$

$$F_1 = 33.1 \pm 0.8$$

$$F_2 = 30.9 \pm 0.8$$

$G$  的主效应和交互作用  $B \times G$  是显著的。这时要对  $B$  的每一个水平进行与  $G_1$ 、 $G_2$  的比较。

$$\overline{B_1 G_1} = 30 + \frac{56}{8} = 37.0 \pm 1.1 \pm 5.49$$

$$\overline{B_1 G_2} = 30 + \frac{41}{8} = 35.1 \pm 1.1 \pm 5.49$$

$$\overline{B_2 G_1} = 30 + \frac{12}{8} = 31.5 \pm 1.1 \pm 5.49$$

$$\overline{B_2 G_2} = 30 + \frac{-45}{8} = 24.4 \pm 1.1 \pm 5.49$$

置信界限前的数字  $\pm 1.1$ , 是为了对  $B_1$  进行与  $\overline{G_1}$ 、 $\overline{G_2}$  的比较, 对  $B_2$  进行与  $\overline{G_1}$ 、 $\overline{G_2}$  的比较。

然而,  $\overline{B_1 G_1}$  和  $\overline{B_2 G_1}$ 、 $\overline{B_1 G_2}$  和  $\overline{B_2 G_2}$  在因素  $B$  的水平不同时, 按下述方法求置信界限。

如果对  $\overline{B_1 G_1}$  进行分解:

$$\overline{B_1 G_1} = \bar{T} + (\overline{B_1} - \bar{T}) + (\overline{G_1} - \bar{T}) + (\overline{B_1 G_1} - \overline{G_1} - \overline{B_1} + \bar{T})$$

这个值所具有的误差, 是一般平均值  $T$ 、 $B$  的主效应  $B_1$  的分量效应、 $G$  的主效应  $G_1$  的分量效应,  $B$  与  $G$  交互作用的第一分量  $G_1 B_1$  的效应之总和。一般平均与  $B$  的主效应是具有一次误差方差  $V_{e_1}$  的估计值。 $G$  的主效应和交互作用  $B \times G$  是具有二次误差方差的估计

值, 因此置信界限可用下式计算 (安全一侧的置信界限):

$$\pm\sqrt{F \times \frac{V_{e_1}}{(n_e)_1} + F \times \frac{V_{e_2}}{(n_e)_2}}$$

上式平方根内的前一个  $F$  值是分子自由度为 1, 分母自由度为 2 的  $F$  表在 5% 点的值等于 18.51,  $(n_e)_1$  是  $e_1$  用检验为显著因素的  $n_1$  与  $B$  的自由度之和除以实验次数 32 所得。即

$$\frac{1}{(n_e)_1} = \frac{1(m) + 1(B)}{32} = \frac{2}{32} = \frac{1}{16}$$

后一个  $F$  是分子自由度为 1, 分母自由度为 21 的  $F$  分布在 5% 点的值,  $(n_e)_2$  是用  $e_2$  检验的显著因素  $G$ 、 $B \times G$  的自由度之和除以实验次数 32 求得, 即

$$\frac{1}{(n_e)_2} = \frac{1(G) + 1(B \times G)}{32} = \frac{2}{32} = \frac{1}{16}$$

置信界限为

$$= \sqrt{18.51 \times \frac{25.5}{16} + 4.32 \times \frac{2.3}{16}} = \pm 5.49$$

它可用于  $\overline{B_1 G_1} = 37.0$  与  $\overline{B_2 G_1} = 31.5$  的比较, 它们的差  $37.0 - 31.5 = 5.5$  要比 5.49 的  $\sqrt{2}$  倍, 即 7.76 小, 所以  $\overline{B_1 G_1}$ 、 $\overline{B_2 G_1}$  之间无显著差异 (见图 46.2-1)。

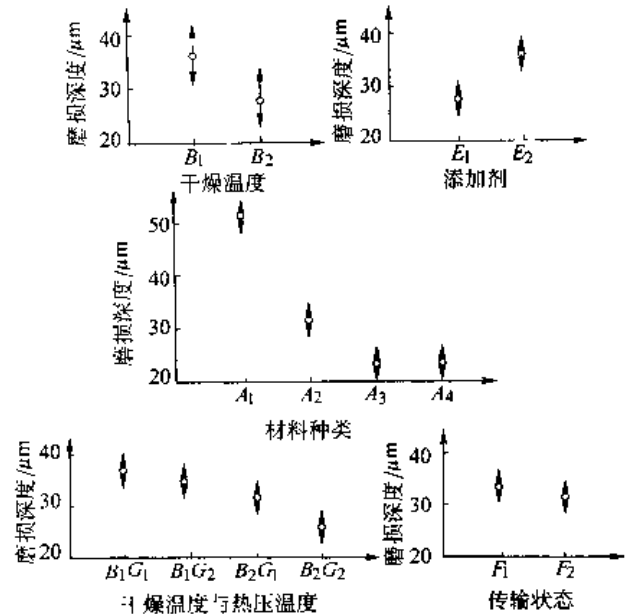


图 46.2-1 因素效应估计

在  $A_3 B_2 G_2 E_1 F_2$  最宜条件的情况下, 工序平均值的估计值为

$$\hat{\mu} = \bar{A}_3 + \overline{B_2 G_2} + \bar{E}_1 + \bar{F}_2 - 3\bar{T}$$

$$= 22.1 + 24.4 + 27.8 + 30.9 - 3 \times 32.0$$

$$= 9.2$$

因为存在一次误差  $e_1$ , 二次误差  $e_2$ , 所以在最条件下估计验证试验的各个值的范围是麻烦的。

$$x = \hat{\mu} \pm \sqrt{F \times V_{r_1} \left( \frac{1}{(n_1)_1} + \frac{1}{4r} \right) + F \times V_{r_2} \left( \frac{1}{(n_2)_2} + 1 - \frac{1}{4r} \right)}$$

$$= 9.2 \pm \sqrt{18.51 \times 25.5 \left( \frac{1}{16} + \frac{1}{8} \right) + 4.32 \times 2.3 \left( \frac{7}{32} + \frac{7}{8} \right)}$$

$$= 9.2 \pm 9.9$$

置信界限为 9.9, 这是因为一次误差方差的自由度较小的缘故。对于各个  $x$  的协方差以及  $\hat{\mu}$  的有效反复数有下列公式计算:

$$V_{ar} = \frac{1}{8} (\sigma_2^2 + \sigma_1^2) + \frac{7}{8} \sigma_2^2$$

$$- \frac{1}{8} V_{r_1} + \left( 1 - \frac{1}{8} \right) V_{r_2}$$

$$\frac{1}{(n_2)_1} = \frac{1(m) + 1(B)}{32} = \frac{2}{32}$$

$$\frac{1}{(n_2)_2} = \frac{3(A) + 1(E) + 1(F) + 1(G) + 1(B \times G)}{32}$$

$$= \frac{17}{32}$$

## 1.4 三水平系统的裂区设计

(话筒用碳粒实验)

### 1.4.1 因素与水平

电话的原理是在话筒中放入碳粒, 当声音在振膜上产生的压力变化时, 使碳粒接触点的电阻变化。然而为了声音的传输尽可能的保真, 要求碳粒具有如下的性质:

- 1) 表面要光滑, 要具有粒子的流动性;
- 2) 电阻、比重等物理量的可变范围要大;
- 3) 吸湿率小;
- 4) 硬度要大, 不易碎且稳定。

此外还要容易制造, 成本低。将碳氢化合物的适当聚合物经过处理后, 代替天然的无烟煤为原料制造碳粒, 具有上述 1)~3) 的性质。所以碳粒实验就是使用这一方法, 对球状的共聚物进行各种处理, 测量它的灵敏度、电阻、吸湿率、比重等特性指标, 以便找出优良的碳粒制造方法。

工序大致如下:



在上述工序中原料的配制与聚合是十分复杂的, 在此跳过这两个工序, 对粒子筛选以后的工序进行实验。设因素如下:

- A: 粒子大小: 40~50, 50~70 (目)  
 B: 氧化升温速度: 25, 50, 100 (°C/h)  
 C: 最后氧化温度在 250°C 的保持时间: 2, 12, 6, 80 (h)

这种聚合物如直接进行高温处理, 大部分会由于

氧化而损失, 收率很低, 只有百分之几, 所以需要采用必要的措施。预计酸化的时间越长, 收率越高。

D: 碳化升温速度: 25, 50, 100 (°C/h)

经过氧化处理以后如直接放入非氧化气体中进行高温处理, 当加温到 600~700°C 时, 将会产生各种挥发物。为防止烧成时被挥发物所污染, 在最终处理即烧成之前, 将产品放到氮气中去, 在 700°C 温度条件下进行 1h 的热处理, 即为碳化工序。

E: 烧成气体: N<sub>2</sub>、H<sub>2</sub>

F: 烧成气体的流量, 见表 46.2-6

表 46.2-6 烧成气体的流量 (cm<sup>3</sup>/min)

	小流量 (F <sub>1</sub> )	大流量 (F <sub>2</sub> )
N <sub>2</sub>	400	800
H <sub>2</sub>	800	1600

G: 最终烧成温度: 1000, 1100, 1200 (°C)

H: 烧成温度升温速度: 25, 50, 100 (°C/h)

烧成条件是最重要的, 所以另外将镍容器中的三个不同部位取做区组因素。

R: 容器中的三个不同部位: 上、中、下

设烧成温度以 25°C/h 的速度上升到 1200°C。因为即使开始的温度已接近碳化程度, 还得需要 20h, 所以 E、F、G、H 有关的烧成因素作为一次因素。并且按进行 9 次实质性实验进行配列。

### 1.4.2 配列与数据

假设因素为:

一次因素 E, F, G, H, 2<sup>2</sup> × 3<sup>2</sup> 型

二次因素 A, B, C, D, R 3<sup>4</sup> × 2<sup>1</sup> 型

配列于正交表 L<sub>27</sub> (3<sup>3</sup>) 中 (见表 46.2-7), 其中在二水平的因素中, 安排如下的拟水平:

A<sub>3</sub> = A<sub>1</sub> = 40~50 (目), E<sub>3</sub> = E<sub>1</sub> = N<sub>2</sub>, F<sub>3</sub> = F<sub>1</sub> = 小流量

其测量值有四种 (见表 46.2-8)。

1) 灵敏度: 四号话筒的振动器测试后经计算所得效率  $\eta$  作为灵敏度

$$\eta = 20 \lg \frac{e}{kV}$$

式中的  $e$  是碳粉交流输出电压,  $k$  是电极的位移,  $V$  是直流电压。

2) 静电阻: 是用一定量的处于静止的碳粒通过 50mA 的电流时的直流电阻值。

3) 吸湿率: 在 100°C 温度下, 干燥了 1h 之后约 0.5cm<sup>3</sup> 的碳粒 W<sub>0</sub> 克; 在温度 30°C, 湿度 80% 的密闭容器中放置 4h 后测得质量 W<sub>1</sub>, 用下式表示吸湿率

$$\frac{W_1 - W_0}{W_0} \times 100\%$$

表 46.2-7 实验配列

因素列 No.	一 次 因 素				二 次 因 素								
	G 1	H 2	E 3	F 4	D 5	G × D 6	D 7	C 8	G × C 9	C 10	B 11	A 12	R 13
1					1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3					3	3	3	3	3	3	3	3	3
4					1	1	1	2	2	2	3	1'	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6					3	3	3	1	1	1	2	2	2
7					1	1	1	3	3	3	2	2	2
8	1	3	1'	1'	2	2	2	1	1	1	3	1'	3
9					3	3	3	2	2	2	1	1	1
10					1	2	3	1	2	3	1	2	3
11	2	1	2	1'	2	3	1	2	3	1	2	1'	1
12					3	1	2	3	1	2	3	1	2
13					1	2	3	2	3	1	3	1	2
14	2	2	1'	1	2	3	1	3	1	2	1	2	3
15					3	1	2	1	2	3	2	1'	1
16					1	2	3	3	1	2	2	1'	1
17	2	3	1	2	2	3	1	1	2	3	1	1	2
18					3	1	2	2	3	1	1	2	3
19					1	3	2	1	3	2	1	1'	2
20	3	1	1'	2	2	1	3	2	1	3	2	1	3
21					3	2	1	3	2	1	3	2	1
22					1	3	2	2	1	3	3	2	1
23	3	2	1	1'	2	1	3	3	2	1	1	1'	2
24					3	2	1	1	3	2	2	1	3
25					1	3	2	3	2	1	2	1	3
26	3	3	2	1	2	1	3	1	3	2	3	2	1
27					3	2	1	2	1	3	1	1'	2
	烧成温度	烧成升温速度	烧成气体	流量	碳化升温速度			氧化保持时间			氧化升温速度	粒子的大小	容器的位置上、中、下

表 46.2-8 实验数据

实验号	效率 /dB	静电阻 /Ω	吸湿率 (%)	容量 /g·cm <sup>-3</sup>
1	89.9	118	9.39	0.728
2	90.6	214	17.68	0.634
3	90.4	215	13.93	0.590
4	90.2	175	20.98	0.560
5	90.1	198	18.22	0.538
6	88.6	112	16.29	0.741
7	90.0	288	14.38	0.668
8	89.2	111	16.98	0.742
9	90.7	299	13.66	0.581
10	88.6	47	6.23	0.754
11	89.6	42	4.78	0.502
12	89.8	43	3.90	0.559
13	90.6	63	3.54	0.638
14	89.9	100	3.84	0.657
15	89.0	55	7.62	0.724
16	90.4	55	2.50	0.568
17	90.0	55	5.22	0.754
18	89.6	85	2.24	0.661
19	91.0	76	12.39	0.736
20	91.2	131	6.90	0.515
21	91.2	92	7.83	0.698
22	89.2	123	10.12	0.693
23	90.4	166	10.74	0.607
24	89.5	87	13.64	0.733
25	89.9	66	10.38	0.597
26	87.6	61	12.84	0.768
27	90.7	79	10.64	0.547
平均	89.9	117	10.25	0.649
比较样品	90.0			

注：比较样品是把无烟煤在 1090°C 的湿度下，在氮气中经 3h 烧成的。其收率由于某种原因不予发表。

4) 容量：在测量静电阻和灵敏度时所用的测量碳粉容积规格为 0.35mm 的量筒里，最紧密地装入 0.5cm<sup>3</sup> 的碳粉，其质量的两倍值表示碳粉的容量，单位 g/cm<sup>3</sup>。

1.4.3 数据分析

在本实验中，二水平因素有 3 个，三水平因素有 6 个。因此对实验所要假定的未知数，即自由度如下：

- (1) 全平均  $m$  为 1 个；
- (2) 三个二水平因素的二水平之间的差值为 3 个；
- (3) 六个三水平因素的主效应有 12 个；
- (4) 交互作用有 8 个。

于是要考虑共有 24 个未知数的方程。例如，要知道烧成气体 N<sub>2</sub> 和 H<sub>2</sub> 对效率产生什么样的效应，就要计算对于 N<sub>2</sub>、H<sub>2</sub> 的平均的差值；用 N<sub>2</sub> 进行 18 次实验（是烧成气体因素 E 所对应的列与 1 实验）的  $\eta$  值的总和除以 18，用 H<sub>2</sub> 进行 9 次实验（烧成气体因素 E 所对应的列与 2 实验） $\eta$  值的总和除以 9，而两者的差就是按最小二乘法求得气体 N<sub>2</sub> 和 H<sub>2</sub> 的差。

对于本例：

N<sub>2</sub> 的平均值

$$E_1 = 90.0 + \frac{1}{18}(-0.1 + 0.6 + 0.4 + 0.0 - 0.8$$

$$\begin{aligned}
 &+0.7+0.6-0.1+0.4+0.0-0.4 \\
 &+1.0+1.2+1.2-0.8+0.4-0.5) \\
 &=90.0+\frac{1}{18}(3.8)
 \end{aligned}$$

H<sub>2</sub> 的平均值

$$\begin{aligned}
 E_2 &=90.0+\frac{1}{9}(0.2+0.1-1.4-1.4 \\
 &\quad -0.4-0.2-0.1-2.4-0.7) \\
 &=90.0+\frac{1}{9}(-4.9)
 \end{aligned}$$

因此 N<sub>2</sub> 比 H<sub>2</sub> 增加了 90.21-89.45=0.76dB 的灵敏度。这一值的正确度如何可用定量地评定 E, G, H, F, D, G×D, C, G×C, B, A, R 以外的因素所产生的影响有多大的误差方差估计值来判断。具体是用方差分析表和估计值的置信界限来进行。

S<sub>E</sub> 可用下式计算, 因为 E<sub>1</sub> 中包含了拟水平。

$$\begin{aligned}
 S_E &= \frac{[N_2\text{-之和}-2(H_2)\text{-之和}]^2}{1 \times 18 + (-2)^2 \times 9} \\
 &= \frac{[3.8-2(-4.9)]^2}{54} \\
 &= 3.42 \quad (f=1)
 \end{aligned}$$

方差分析表(见表 46.2-9)的检验如下:

(1) 二次误差的自由度只有 1, 所以把方差小的因素归并到 e<sub>2</sub>, 使其自由度达到总自由度的一半, 即 9 左

右。把 D, G×C, B 归并到 e<sub>2</sub>, 得 (e<sub>2</sub>)。

(2) 用归并而成的 (e<sub>2</sub>) 的方差 0.10 去检验二次因素和 e<sub>1</sub>。因素 C, A, R 是显著的, e<sub>1</sub> 是不显著的。

(3) 与 e<sub>1</sub> 大致相同的 G, G×D 归并到 e<sub>1</sub>, 组成最终的误差方差 (e)。

(4) 用 (e) 做检验。这时归并而成的因素可以用它做检验。归并的目的只是为了更正确地估计所考虑的误差方差, 如果用自由度小的波动去估计误差方差, 所得的值可能会是真值的 1/10。归并时虽然有误差, 但一般只有百分之几的程度。

(5) 对检验结果为显著的因素, 可以用图 46.2-2 表示。

在没有拟水平时, 置信界限的计算为

$$\pm \sqrt{4.45 \times \frac{0.184}{9}} = \pm 0.30$$

在有拟水平时, 置信界限的值为上述值的  $\frac{1}{\sqrt{2}}$ ,

即 ±0.20。

同样容量的方差分析表和图分别为表 46.2-10 和图 46.2-3 所示。其中 (e<sub>2</sub>) 是把 G×D, G×C, R 归并到 e<sub>2</sub> 的误差方差, (e) 是把 e<sub>1</sub>, G, F 归并到 (e<sub>2</sub>) 后的误差方差。

从表 46.2-9 可知, 因素 H, E, F, C, A 都对效

表 46.2-9 效率的方差分析表

因素	f	S	V	F <sub>0</sub>	S'	p (%)
一次因素	C	0.60	0.30	—	—	—
	H	1.52	0.76	4.1	1.15	6.2
	E	2.94	2.94	16.0	2.76	15.0
	F	1.50	1.50	8.2	1.32	7.2
	一次误差 e <sub>1</sub>	2	0.58	0.29	—	—
二次因素	D	0.08	0.04	—	—	—
	G×D	1.05	0.26	1.4	—	—
	C	5.81	2.90	15.8	5.44	29.6
	G×C	0.32	0.08	—	—	—
	B	0.44	0.22	—	—	—
	A	2.67	2.67	14.5	2.49	13.4
	R	1.00	0.50	2.7	—	—
二次误差 e <sub>2</sub>	1	0.04	0.04	—	7.39	39.3
(e <sub>2</sub> )	(9)	(0.90)	(0.10)	—	—	—
(e)	(17)	(3.13)	(0.184)	—	(5.39)	(23.6)
总计	26	18.55	—	—	18.55	100.0

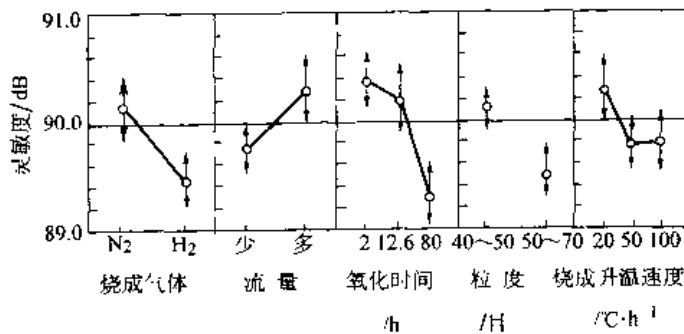


图 46.2-2 效率的因素效应

率产生影响,  $E, C$  的效应的影响率为 20% 左右, 其他因素的和约 39%。另一方面因素  $E, D, C, B, A$  对容量有显著影响, 其中仅氧化时间  $C$  这一因素的影响率就占 2/3, 其他因素的效应只有 6.3%。

由图 46.2-2 可见, 如果要使灵敏度达到最高, 烧成气体用  $N_2$ , 流量最大, 氧化时间 2h, 粒度为 40~50 目, 烧成温度的上升速度要慢些。在这种条件下生产, 预计灵敏度为:

$$\hat{\mu} = (\text{在 } N_2 \text{ 的平均值}) + (\text{在流量最大 } F \text{ 的平均值})$$

$$n_e = \frac{27}{1(m) + \left(\frac{3}{2}-1\right)(N_2) + (3-1)(多) + (3-1)(2h) + \left(\frac{3}{2}-1\right)(40\sim50) + (3-1)}$$

$$= \frac{27}{1+0.5+2+2+0.5+2} = \frac{27}{8}$$

验证实验时所得的各个值  $x$  的预测值为:

$$x = \hat{\mu} \pm \sqrt{F \times V_r \left( \frac{1}{n_e} + 1 \right)}$$

$$= 91.44 \pm \sqrt{4.45 \times 0.184 \left( \frac{8}{27} + 1 \right)}$$

$$= 91.44 \pm 1.03$$

表 46.2-10 容量的方差分析表

因素	$f$	$S$	$V$	$F_0$	$\rho$ (%)	
一次	$C$	2	7.3	3.7	—	—
	$H$	2	22.1	11.1	3.0	—
	$E$	1	117.0	117.0	31.9	6.1
	$F$	1	6.0	6.0	—	—
	$e_1$	2	1.1	0.6	—	—
二次	$D$	2	28.1	14.1	3.8	1.2
	$G \times D$	4	11.5	2.9	—	—
	$C$	2	1214.0	607.0	165.4	66.1
	$G \times C$	4	25.6	6.4	—	—
	$B$	2	57.7	28.9	7.2	2.8
	$A$	1	327.1	327.1	89.1	17.5
	$R$	2	4.7	2.4	—	—
	$e_2$	1	2.5	2.5	—	—
	$(e_2)$	(11)	(44.3)	(4.03)	—	—
$(e)$	(16)	(58.7)	(3.67)	—	6.3	
总计	26	1824.7	—	—	100.0	

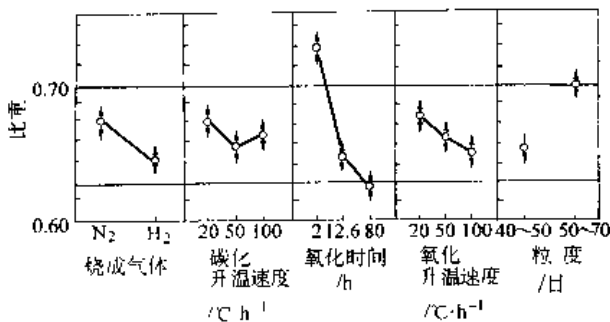


图 46.2-3 容量的因素效应

### 1.5 双元裂区设计

二方分割法 (铜的电解实验)

值) + (在氧化时间 2h 下的平均值) + (在粒度 40~50 目下的平均值) + (在烧成升温速度 20°C/h 下的平均值) - 4 (总平均)

$$90.14 + 90.27 + 90.36 + 90.10$$

$$+ 90.25 - 4 \times 89.91 = 91.44$$

91.44 是实验中未出现的最高值。在这个条件下, 可以简单算出任何一个其他特性值。

这时由于烧成气体、流量和粒度中安排有拟水平, 所以  $n_e$  由下式计算:

27

在铜的电解实验中有四个实验用的电解槽, 各个槽中的电解溶液依靠一台泵进行循环流动, 因此电解溶液的成分, 在四个槽中不一样时就不能进行实验。阳极是粗铜, 但每制成一根电解铜棒, 不要把阳极上变细的粗铜棒取下来, 直到完全消耗完, 即四根电解铜棒得到之前不要去动它。用电解法制得一根铜棒大约需要一星期。现选用如表 46.2-11 所示的因素。

表 46.2-11 对因素、水平数以及反复试验次数的要求

	因素	水平数	时间要求	槽的要求
与电解液有关的	$A$	2	可以每周变一次	四个槽都必须是相同的
	$B$	4		
	$C$	2		
	$D$	2		
	$E$	2		
	$F$	2		
	$G$	2		
	$H$	2		
	$I$	4		
	$J$	2		
与阳极有关的	$K$	2	只要每四周更换一次	每个槽不一样也可
	$L$	2		
	$M$	2		
	$N$	2		
	$O$	2		
与槽有关的	$R$ (4 槽)	( $W$ ) 4	实验中保持同一条件, 只要每四周更换一次	每个槽可以更换
	$P$			
	$Q$			
与阴极有关的	$S$	2	每周可以更换一次	每个槽可以被更换
	$T$	2		
	$U$	2		
	$V$	4		

要求全部因素的主效应和交互作用  $K \times L, L \times B, K \times B, K \times L \times B, C \times D$  请见配列。用图 46.2-4 解释表 46.2-11。

三个不同单位的配列很难进行,利用点线图进行帮助配列。现采用正交表  $L_{64}$  型 (5)。正交表  $L_{64}$  型 (5) 的 5 个五边形有共同的底边,但各个五边形在  $L_{64}$  中的分别构成  $L_{16}$ 。因此若使  $K \sim Q$  的因素与一个五边形相对应,使  $A \sim J$  的因素与另一五边形相对应,则剩下的因素可以排入  $L_{64}$  的其余任意一列。必须使  $K \sim Q$  的五边形与  $A \sim J$  的五边形有一条公共线,把四周作为一个区,使这一线段的四个水平对应于一个四水平的因素。

这时要使与电解液有关的因素与上方的五边形相对应,必须利用与  $L_{16}$  的对应关系改变五边形的点线图。这样得到点线图 (见图 46.2-5)。

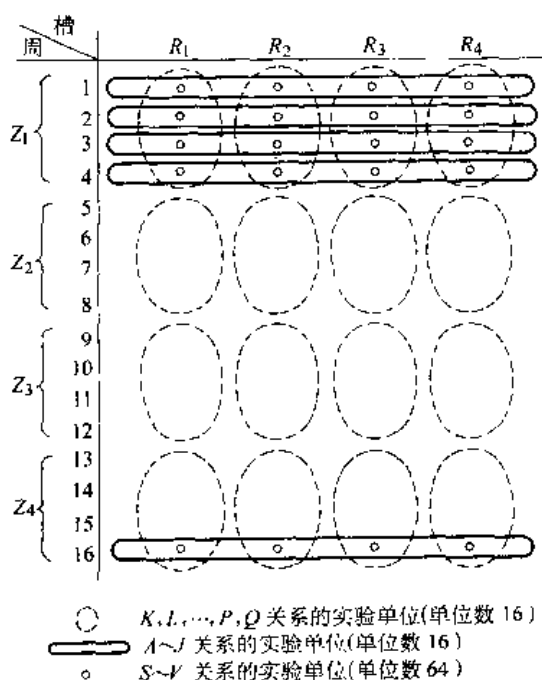


图 46.2-4 二方分割法的示意图

除第 1~3 列外,右上方五边形线段与正交表  $L_{16}$  点线图的线段对应如下:

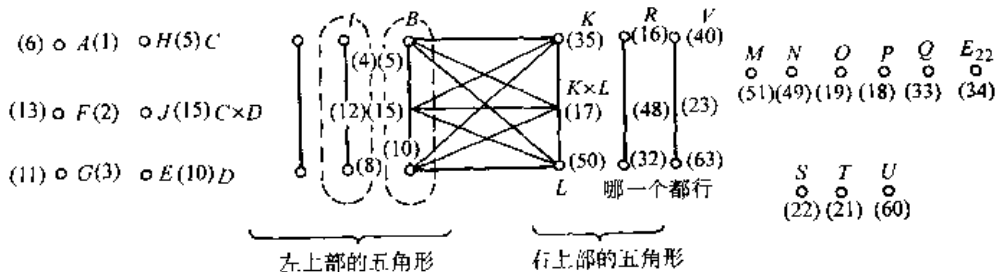


图 46.2-5 所要求的点线图

## 2 拟因素设计

当各种水平的因素混杂在一起或者若干产品一起进行实验时,为解决数据的复杂性所使用的方法就是

32-4, 49-10, 33-5, 48-11, 34-6, 19-12, 35-7, 18-13, 51-8, 17-14, 50-9, 16-15

利用这种对应,如引用型  $L_{16}$  (5) 就可能按图中括号内进行配列。从交互作用表可知,  $B \times K$  是 38, 41, 44,  $B \times L$  是 55, 56, 61,  $B \times K \times L$  是 20, 27, 30。剩下的构成三次误差  $e_3$ 。表 46.2-12 所示为因素分析表。

表 46.2-12 使用二方分割法时的方差分析表

因素	f	列	E(V)
H	1	1	$\sigma_3^2 + 4(\sigma_{21}^2 + \sigma_{22}^2) + 16\sigma_1^2 + 32\sigma_2^2$
J	1	2	$\sigma_3^2 + 4(\sigma_{21}^2 + \sigma_{22}^2) + 16\sigma_1^2 + 32\sigma_2^2$
E	1	3	$\sigma_3^2 + 4(\sigma_{21}^2 + \sigma_{22}^2) + 16\sigma_1^2 - 32\sigma_2^2$
$e_1$	0	-	$\sigma_3^2 + 4(\sigma_{21}^2 + \sigma_{22}^2) + 16\sigma_1^2$
C	1	7	$\sigma_3^2 + 4\sigma_{21}^2 + 32\sigma_2^2$
D	1	9	$\sigma_3^2 + 4\sigma_{21}^2 + 32\sigma_2^2$
$C \times D$	1	14	$\sigma_3^2 + 4\sigma_{21}^2 - 16\sigma_{C \times D}^2$
I	3	4, 8, 12	$\sigma_3^2 + 4\sigma_{21}^2 + 16\sigma_7^2$
B	3	5, 10, 15	$\sigma_3^2 + 4\sigma_{21}^2 + 16\sigma_8^2$
A	1	6	$\sigma_3^2 + 4\sigma_{21}^2 + 32\sigma_3^2$
F	1	13	$\sigma_3^2 + 4\sigma_{21}^2 + 32\sigma_7^2$
G	1	11	$\sigma_3^2 + 4\sigma_{21}^2 + 32\sigma_8^2$
$e_{21}$	0	-	$\sigma_3^2 - 4\sigma_{21}^2$
R	3	16, 32, 48	$\sigma_3^2 + 4\sigma_{22}^2 + 16\sigma_{R(W)}^2$
K	1	35	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_K^2$
L	1	50	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_L^2$
$K \times L$	1	17	$\sigma_3^2 + 4\sigma_{22}^2 + 16\sigma_{K \times L}^2$
M	1	51	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_M^2$
N	1	49	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_N^2$
O	1	19	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_O^2$
P	1	18	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_P^2$
Q	1	33	$\sigma_3^2 + 4\sigma_{22}^2 + 32\sigma_Q^2$
$e_{22}$	1	34	$\sigma_3^2 + 4\sigma_{22}^2$
V	3	40, 23, 63	$\sigma_3^2 + 16\sigma_V^2$
S	1	22	$\sigma_3^2 + 32\sigma_S^2$
T	1	21	$\sigma_3^2 + 32\sigma_T^2$
U	1	60	$\sigma_3^2 + 32\sigma_U^2$
$B \times K$	3	38, 41, 44	$\sigma_3^2 + 16\sigma_{B \times K}^2$
$B \times L$	3	55, 56, 61	$\sigma_3^2 + 16\sigma_{B \times L}^2$
$B \times K \times L$	3	20, 27, 30	$\sigma_3^2 + 8\sigma_{B \times K \times L}^2$
$e_3$	12	全部剩余	$\sigma_3^2$
总计	63		

拟因素法。这种方法是实验设计中最重要方法。

### 2.1 什么是拟因素法

拟因素法是把各种复杂的因素配列到一张正交表



中去的一种有效方法。拟因素法有下列两种方法组成：一种是按某个因素的水平，对不同的因素进行试验的变身法；另一种是把三水平、五水平、七水平等因素配列于二水平的正交表中去，七水平因素配列于三水平正交表中去的赋闲列法。

并且，当一个或两个以上因素已指定了区组大小的区组因素时，也使用拟因素法进行巧妙设计。

### 2.2 拟因素设计中的改造因素技术

以碳膜电阻实验为例。

#### 2.2.1 配列与数据

碳膜电阻是用适当的工艺使碳粉喷覆在陶瓷表面而制成的。假定取两个因素，陶瓷种类作为因素  $A$ ，取二水平；喷覆工序的温度作为因素  $B$ ，取二水平。喷覆碳粉后有  $C_1$ 、 $C_2$  两种加工方法：

$C_1$  是现行的涂覆瓷漆方法； $C_2$  是模压酚醛树脂的新方法。现在要比较哪一种方法较好。已知涂覆瓷漆法和模压酚醛树脂法的最宜条件与二水平因素  $A$ 、 $B$  一起进行实验。现对瓷漆涂覆法取下列两个因素

- $D'$ ：瓷漆种类 二水平
- $E'$ ：涂覆次数 二水平

找到了它的最宜条件后再与模压酚醛树脂法的最宜条件作比较。对于也取  $D''$  作为因素：

- $D''$ ：模压温度 二水平

比较  $D'_1$ 、 $D''_2$  的好坏并将好的水平与瓷漆涂覆最宜条件进行比较。

现在的问题就是求出二水平因素  $A$ 、 $B$  的主效应， $C_1$  中的  $D'$ 、 $E'$  的主效应， $C_2$  中  $D''$  的主效应并找出  $C_1$ 、 $C_2$  的最宜条件后，比较  $C_1$ 、 $C_2$ 。这样对  $C_1$  或  $C_2$  研究不同的因素时，可作如下配列（见表 46.2-13）：

表 46.2-13 自由度的分配

因素	$f$
$C$	1
$D$	1
$C \times D$	1
总计	3

当求因素  $C$ 、 $D$  的交互作用时，通常对数据做如表 46.2-13 所示的分解，这样分解是为了使交互作用变小或是为零。如果  $D$  的效应对  $C_1$  与  $C_2$  有很大的差异时，就必须对  $C_1$ 、 $C_2$  求  $D$  的效应，即必须对数据作如表 46.2-14 所示的分解。

不管  $D$  的效应对  $C_1$  与  $C_2$  有怎样的差异，只要分别求出  $C_1$  与  $C_2$  时的  $D$  的效应。由于进行上述的分解  $D$  的效应对  $C_1$  与  $C_2$  有怎样的差异都没有关系，所以  $D$  也是在  $C_1$ 、 $C_2$  时的不同因素，也就是说因素  $D$  既是

对  $C_1$  为瓷漆种类的二水平因素  $D$ ，又是对  $C_2$  变为模压温度的因素  $D$ ，对应作如下的分解（见表 46.2-15）。

表 46.2-14 自由度的分配

因素	$f$
$C$	1
$D$ (对应 $C_1$ )	1
$D$ (对应 $C_2$ )	1
总计	3

表 46.2-15 自由度的分配

因素	$f$
$C$	1
$D'$ (对应 $C_1$ )	1
$D''$ (对应 $C_2$ )	1
总计	3

这种对应  $C_1$  时是  $D'$ ，对应  $C_2$  时变为  $D''$  的因素  $D$ ，称为拟因素。即在求  $C_1$ 、 $C_2$  不同因素的主效应时，要在形式上保证因素  $C$  和因素  $D$  的交互作用。（见图 46.2-6）。

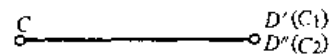


图 46.2-6 变身法的点线图

同样，如果只对  $C_1$  求  $E'$  因素的主效应，就要保证因素  $C$  与拟因素  $E$  的交互作用，对应  $C_1$  时  $E$  为  $E'$ ，在对应  $C_2$  时  $E$  就变为误差。

此时因素的配列可以用图 46.2-7 同正交表  $L_8$  对应起来的是表 46.2-16 所示的配列。由于去掉反映第 1 列（对应  $C$ ）与第 2 列（对应拟因素  $D$ ）的交互作用的第 3 列，所以当第 1 列是  $C_1$  和  $C_2$  时，就使用第 2 列。由于去掉反映第 1 列与第 4 列的交互作用的第 5 列，所以当第 1 列只是一个水平时，就把因素  $E$  与第 5 列相对应；当第 1 列是两个水平时，第 5 列就变成实验误差。

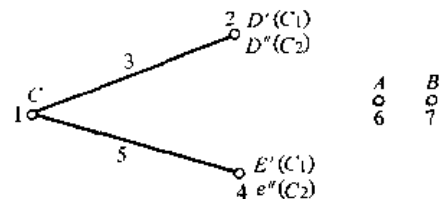


图 46.2-7 碳膜电阻实验的点线图

表中的数据是在某种强制实验条件下的碳膜电阻值的变换率，是由原数据减去虚拟平均值再扩大 10 倍而得。

第三个实验室在  $A_2B_2$  条件下制造两个碳膜电阻，此时  $C$  就是  $C_1$ ，所以在  $D'_2E'_1$  的条件下用瓷漆涂覆法加工。如果作寿命实验则如表所示其中一个碳膜的变化率为 6.2%，另一个为 6.4%。第一~四个实验是瓷漆涂覆，第五~八个是模压酚醛树脂。

表 46.2-16 碳膜电阻实验的配列和数据

因素 列 No	D'		E'		数据	(y - 4.0) ×10		总计
	C	D''	e	A B				
	1	2	3	4	5	6	7	
		D'		E'				
1	1	1	1	1	1	3.0	3.2	-10 -8 18
2	1	1	2	2	2	3.5	3.3	-5 -7 -12
3	1	2	1	2	2	6.2	6.1	22 24 46
4	1	2	2	1	1	5.1	5.2	11 12 23
		D''		e <sub>1</sub>				
5	2	1	1	1	2	3.8	3.5	-2 -5 7
6	2	1	2	2	1	3.0	2.8	-10 -12 -22
7	2	2	1	2	1	3.5	3.3	-5 -7 -12
8	2	2	2	1	2	4.4	4.6	4 6 10
总 计								8

2.2.2 方差分析

在方差分析中，波动值的计算如下：

$$S_m = \frac{(40 \times 16 + 8)^2}{16} = 25600 + 640 + \frac{8^2}{16} = 26244 \quad (f=1)$$

对于诸如变化率、磨损量等数据，它们的目标值为零，从而一般平均值偏离零的多少是一个重要情报，因此最好求出。注意此时应当不能减去虚拟平均值的  $S_e$ 。

$$CF = \frac{8^2}{16} = 4$$

$$S_{T_1} = [(-18)^2 + (-12)^2 + (46)^2 + (23)^2 + (-7)^2 + (-22)^2 + (-12)^2 + (10)^2] = 1941 \quad (f=7)$$

$$S_A = \frac{(8-0)^2}{16} = 4 \quad (f=1)$$

$$S_B = \frac{(-29-37)^2}{16} = 272 \quad (f=1)$$

C 的效应首先要对整个瓷漆涂覆和模压酚醛树脂作比较。

$$S_C = \frac{(39+31)^2}{16} = 306 \quad (f=1)$$

其次瓷漆涂覆因素 D', E' 的效应，可以根据 C<sub>1</sub> 的第一~四个实验求得。

$$S_{D'} = \frac{[-18-12-(46+23)]^2}{8} = 1225 \quad (f=1)$$

$$S_{E'} = \frac{(28-11)^2}{8} = 36 \quad (f=1)$$

D'' 的效应可以根据 C 的第五~八个实验求得。

$$S_{D''} = \frac{[-29-(2)]^2}{8} = 91 \quad (f=1)$$

实验误差  $S_{e_1}$  以下列两种不同方法求得，但第一式不能用来验算所以一般不用。

$$S_{e_1} = S_{T_1} - (S_A + S_B + \dots + S_{D''})$$

$$= 1941 - (4 + 272 + \dots + 91)$$

$$= 7 \quad (f=1)$$

$$S_{e_1} = C_1 \text{ 的四列波动} = \frac{[-7-12-(-22+10)]^2}{8}$$

$$= 6 \quad (f=1)$$

验算就可以看看  $S_A, S_B, \dots, S_{D''}, S_{e_1}$  之和是否与  $S_{T_1}$  一致。现在有差值为 1，这时可以认为是四舍五入的误差。

重复间误差的波动，可以从总波动中减去求得；当重复数 2 时，也可以由两值之差的平方和除 2 求得。

由此得方差分析表（见表 46.2-17）。

表 46.2-17 方差分析表

因素	f	S	V	F <sub>0</sub>	S'	ρ (%)
m	1	26244	26244	9720.0	26241.3	93.0
A	1	4	4	—	—	—
B	1	272	272	100.7	269.3	1.0
C	1	306	306	113.3	303.3	1.1
D' (C <sub>1</sub> )	1	1225	1225	453.7	1222.3	4.3
E' (C <sub>1</sub> )	1	36	36	13.3	33.3	0.1
D'' (C <sub>2</sub> )	1	91	91	33.7	88.3	0.3
e <sub>1</sub>	1	7	7	—	—	—
e <sub>2</sub>	8	17	2.1	—	—	—
(e)	(10)	(27)	(2.7)	—	43.2	0.2
总计	16	28202			28202.0	100.0

但是总的波动  $S_T$  还要加上  $S_m$ ：

$$S_T = S_m - S_{T_1} + S_{e_2} = 26244 + 1941 + 17 = 28202 \quad (f=16)$$

把 A, e<sub>1</sub> 并入 e<sub>2</sub> 得 (e)，用它检验的结果，如表 46.2-17 所示。

一般平均值 m 的影响率  $\rho_m$  含义如下。如果  $\rho_m$  较小，且与各因素影响率之和大致相等，就意味着挑选适当的因素水平组合，会有变化率为正、为零、为负的工艺条件。在此例中，m 的影响率比其他大得多时，不管如何选择显著因素的组合，也不能是变化率为零，只能往零靠近。当影响因素的影响率之和为  $\rho_m$  的一半时，就存在变化率为零的条件。当影响因素的影响率之和大于  $\rho_m$  时，就存在许多使变化率为零的条件，应当从中选择成本最低的条件为最宜条件。

2.2.3 显著因素的估计与最宜条件的决定

显著因素的估计

$$T = 4.0 + \frac{0.8}{16} = 4.05 \pm 0.09$$

$$B_1 = 4.0 + \frac{-2.9}{8} = 3.64 \pm 0.13$$

$$B_2 = 4.0 + \frac{3.7}{8} = 4.46 \pm 0.13$$

对于 C, 首先估计整个瓷漆涂覆的一般平均值和整个模压酚醛树脂的一般平均值。

$$\left. \begin{aligned}
 & C \left\{ \begin{aligned}
 C_1 &= 4.0 + \frac{3.9}{8} = 4.49 \pm 0.13 \\
 \bar{C}_1 &= 4.0 + \frac{3.1}{8} = 3.61 \pm 0.13
 \end{aligned} \right. \\
 & D'(C_1) \left\{ \begin{aligned}
 \bar{D}_1 &= 4.0 + \frac{3.0}{4} = 3.25 \pm 0.18 \\
 \bar{D}_2 &= 4.0 + \frac{6.9}{4} = 5.72 \pm 0.18
 \end{aligned} \right. \\
 & E'(C_1) \left\{ \begin{aligned}
 \bar{E}_1 &= 4.0 + \frac{2.80}{4} = 4.70 \pm 0.18 \\
 \bar{E}_2 &= 4.0 + \frac{1.1}{4} = 4.28 \pm 0.18
 \end{aligned} \right. \\
 & D''(C_2) \left\{ \begin{aligned}
 \bar{D}''_1 &= 4.0 + \frac{-2.9}{4} = 3.28 \pm 0.18 \\
 \bar{D}''_2 &= 4.0 + \frac{-0.2}{4} = 3.95 \pm 0.18
 \end{aligned} \right.
 \end{aligned}$$

关于置信界限, 例如  $\bar{T}$  的置信界限计算如下:

$$\pm \sqrt{\frac{4.96 \times 0.027}{16}} = \pm 0.09$$

拟因素法的最重要一点是, 要分别求得瓷漆涂覆法、模压酚醛树脂法的最宜条件后, 还需对两者进行比较。C<sub>1</sub> 的最佳条件是 C<sub>1</sub>D<sub>1</sub>E<sub>2</sub>。因为  $\bar{D}'_1, \bar{D}'_2$  的平均值,  $\bar{E}'_1, \bar{E}'_2$  的平均值都相等, 所以的 C<sub>1</sub>D<sub>1</sub>E<sub>2</sub> 平均值可以由下式求得:

$$\begin{aligned}
 \hat{\mu}(C_1 D_1 E_2) &= \bar{C}_1 + (\bar{D}'_1 - \bar{C}_1) + (\bar{E}'_2 - \bar{C}_1) \\
 &= \bar{D}'_1 + \bar{E}'_2 - \bar{C}_1 \quad (46.2-1) \\
 &= 3.25 + 4.28 - 4.49 \\
 &= 3.04 \pm 0.22
 \end{aligned}$$

置信界限按下式计算:

$$\pm \sqrt{4.96 \times 0.027 \times \frac{6}{16}} = \pm 0.22$$

同样 C<sub>2</sub> 的最宜条件 C<sub>2</sub>D''<sub>1</sub> 平均值为

$$\begin{aligned}
 \hat{\mu}(C_2 D''_1) &= \bar{C}_2 + (\bar{D}''_1 - \bar{C}_2) \\
 &= \bar{D}''_1 = 3.28 \pm 0.18 \quad (46.2-2)
 \end{aligned}$$

此式置信界限计算公式中的 n<sub>e</sub> = 4。

其中

$$\begin{aligned}
 \frac{1}{n_e} &= \frac{1(m \text{ 的主效应}) + 1(C \text{ 的主效应}) + 2(D'_1) + (2E'_2)}{\text{总实验次数}} \\
 &= \frac{6}{16}
 \end{aligned}$$

对于 D<sub>1</sub>, E<sub>1</sub> 是考虑了拟因素 D 和 D×C, 拟因素 E 和 E×C 的估计值, 所以自由度都为 1+1=2。

由式(46.2-1)、式(46.2-2)可知, 在最宜条件下以瓷漆涂覆法为好。3.04 与 3.28 有显著差别, 可将它们的差与两个置信界限的平方和之平方根比较。

$$3.28 - 3.04 = 0.24$$

$$\sqrt{(0.22)^2 + (0.18)^2} = 0.28$$

本例不存在显著差别。

整体最宜条件是 B<sub>1</sub>C<sub>1</sub>D<sub>1</sub>E<sub>2</sub> 或 B<sub>1</sub>C<sub>2</sub>D''<sub>1</sub>, B<sub>1</sub>C<sub>1</sub>D<sub>1</sub>E<sub>2</sub> 的工序平均为  $\hat{\mu}$  为:

$$\begin{aligned}
 \hat{\mu} &= \bar{T} + (B_1 - \bar{T}) + (C_1 D_1 E_2 \text{ 的平均} - \bar{T}) \\
 &= B_1 + \bar{C}_1 \bar{D}'_1 \bar{E}'_2 - \bar{T} \\
 &= 3.64 + 3.04 - 4.05 = 2.63
 \end{aligned}$$

在 B<sub>1</sub>C<sub>1</sub>D<sub>1</sub>E<sub>2</sub> 条件下, 制造产品时各个 x 值的存在范围:

$$\begin{aligned}
 x &= \hat{\mu} \pm \sqrt{F \times v_e \left( \frac{1}{n_e} + 1 \right)} \\
 &= 2.63 \pm \sqrt{4.96 \times 0.027 \left( \frac{7}{16} + 1 \right)} \\
 &= 2.63 \pm 0.44
 \end{aligned}$$

其中显著因素是 m, B, C, D', E', 根据这些因素的有效自由度的和, 得

$$\frac{1}{n_e} = \frac{1(m) + 1(B) + 1(C) + 2(D') + (2E')}{16} = \frac{7}{16}$$

在最宜条件下生产的产品中, 平均 95% 位于的范围内。当然只对一个样品进行实验时, 这个值表明以 95% 的置信度位于 x 范围内取值。

### 2.3 拟因素设计中的赋闲列技术

就是把三水平因素配列到二水平正交表中去的方法。

现有因素 B 的三个水平 B<sub>1</sub>, B<sub>2</sub>, B<sub>3</sub>。拟因素 B' 是对 A<sub>1</sub> 时为 B<sub>1</sub>, B<sub>2</sub>, 对 A<sub>2</sub> 时为 B<sub>2</sub>, B<sub>3</sub> 的二水平因素。如果在正交表中去掉因素 A 与拟因素 B' 的交互作用 A×B' 列, 那么就能在 A<sub>1</sub>, A<sub>2</sub> 的各水平下求出的主效应。也就是说, 在 A<sub>1</sub> 时求得 B<sub>1</sub>, B<sub>2</sub> 之差, 在 A<sub>2</sub> 时求得 B<sub>2</sub>, B<sub>3</sub> 之差, 即自由度按表 46.2-18 分配:

表 46.2-18 自由度的分配

因素	f
A	1
B <sub>1</sub> B <sub>2</sub> (在 A <sub>1</sub> 时)	1
B <sub>2</sub> B <sub>3</sub> (在 A <sub>2</sub> 时)	1
计	3

这时, 因素 A 的主效应 A<sub>1</sub> 与 A<sub>2</sub> 的差中就混杂有 B<sub>1</sub> 与 B<sub>2</sub> 之差, A<sub>2</sub> 时的差, 所以 B<sub>2</sub>, B<sub>3</sub> 之差也能求得, 可以由 A 的效应去修正 B 的效应。但是, 最好尽可能避免这种修正计算。因此如有多余的列, 通常 A 所对应的列中不对应任何因素而作为空列, 并将与空列的交互作用去掉。

这样的空列称为赋闲列。当然没有多余列, 赋闲列可以是对应因素, 但尽可能对应于诸如求出效应也不起作用的区组因素之类的因素。现设 A, B 都是三水平

的, C, D 是二水平的。按赋闲列法设计的点线图如图 46.2-8 所示。

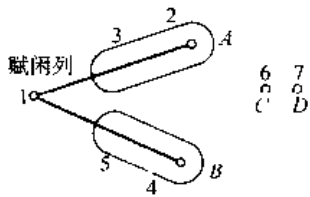


图 46.2-8 赋闲列点线图

此图与正交表对应的表如表 46.2-19 所示。

表 46.2-19 按赋闲列的设计

因素列 No.	赋闲列 1	A 23	B 45	C 6	D 7
1	1	1	1	1	1
2	1	1	2	2	2
3	1	2	1	2	2
4	1	2	2	1	1
5	2	2	2	1	2
6	2	2	3	2	1
7	2	3	2	2	1
8	2	3	3	1	2

自由度按表 46.2-20 与表 46.2-21 分配。

表 46.2-20 自由度的分配

因素	f
赋闲列	1
	1
	1
	1
	1
	1
	0
总计	7

表 46.2-21 自由度的分配

因素	f
赋闲列	1
A <sub>1</sub> A <sub>2</sub>	1
A <sub>2</sub> A <sub>3</sub>	1
B <sub>1</sub> B <sub>2</sub>	1
B <sub>2</sub> B <sub>3</sub>	1
C	1
D	1
e	0
总计	7

## 2.4 发射机碳粉实验

### 2.4.1 因素与水平

用于电话机的碳粒是用无烟煤处理后制成的。现从原料、破碎、烧成工序中取四个因素,以求它们的主效应:

- A: 无烟煤的牌号      二水平
- B: 粒度分布            二水平
- C: 烧成气体            三水平

D: 烧成温度            三水平

在烧成前要对碳粒进行清洗,清洗方法 E

E<sub>1</sub>: 水洗      E<sub>2</sub>: 酸洗

对两种方法进行比较,如果知道了 E<sub>1</sub>、E<sub>2</sub> 的最宜条件,则在最宜条件下进行比较。所以 E 是二水平的,当最宜条件未知时,适当地决定 E<sub>1</sub>、E<sub>2</sub> 两者的比较方法,需要确定对 E<sub>1</sub>、E<sub>2</sub> 分别进行研究的因素,在这些因素水平组合的范围内,求得最宜条件,再做水洗酸洗的比较。

当用同样的制造工艺制造规格性能不同的产品时,也会出现同样的问题。由于分别对普通电容器 E<sub>1</sub> 和飞机用电容器 E<sub>2</sub> 进行实验,实验效率较低,因此将它们一起进行实验。但是 E<sub>2</sub> 有绕线工序而 E<sub>1</sub> 没有。只对 E<sub>2</sub> 绕线工序的若干因素进行研究也一样。

现对水洗 E<sub>1</sub> 取三个因素

F': 水质                  二水平

G': 水温                  二水平

H': 水洗次数            二水平

在求得 F'、G'、H' 的八种组合的最宜条件后,再同 E<sub>2</sub> 的最宜条件进行比较。

下而求酸洗的下列两个因素的主效应

F'': 酸浓度                三水平

G'': 酸洗时间            二水平

并求出六种组合的最宜条件。

结果变成根据实验结果来估计二水平因素 A、B, 三水平因素 C、D 的主效应。在 E<sub>1</sub>、E<sub>2</sub> 的最宜条件下分别求得相应的 E<sub>1</sub>、E<sub>2</sub> 之差,在 E<sub>1</sub> 时 F'、G'、E' 的主效应;在 E<sub>2</sub> 的 F''、G'' 主效应,总共 13 个自由度,即 13 个未知数。自由度为 13 个,所以采用 15 自由度的正交表 L<sub>15</sub> 来配列。按照变身法和赋闲列法来进行配列。

### 2.4.2 配列与数据

用正交表进行实验设计分这三步:

- 1) 把所要求的用点线图表示。
- 2) 一般把标准点线图进行变形,成为要求的点线较,并标上列号。
- 3) 把因素配列于与正交表的列号相对应的列中去。

### 2.4.3 对碳粒实验进行实验设计

(1) 画出所要求的点线图

F'' 是三水平,可以赋闲列法进行配列。由于 F'' 只是与 E<sub>2</sub> 有关的因素,所以 F'' 与 E 的交互列必须去掉。对于 E<sub>1</sub>,可以利用的 F'' 列进行配列。假定在 E<sub>1</sub> 时,安排 F'' 的列安排 F', G', 那么要求的点线图如图 46.2-9 所示。其中三水平的 C、D 共有二个 F'' 的赋闲列,与 E<sub>1</sub> 相关的因素 H' 与因素 G'' 在 E 之间使用变身法。

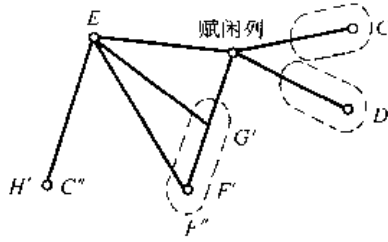


图 46.2.9 要求的点线图

(2) 改造标准的点线图并标上列号

在正交表  $L_{16}$  的点线图中找出与图 46.2-9 最接近的点线图，为 (2) 型点线图。在 (2) 型点线图中，其中新的列即第 1 列与图 46.2-9 的赋闲列对应。把第 1 列作为赋闲列，利用第 15 列和赋闲列与第 15 列的交互作用列即第 14 列配列  $F''$ 。但是  $F''$  是只与  $E_2$  有关的因素，让  $E$  与第 8 列相对应之后，第 8 列与第 15 列、14 列、赋闲列的交互作用的三列可以去掉。这时赋闲列也包含了  $F''$  情报。因此第 8 列与赋闲列的交互作用列即第 9 列也可以去掉。 $E_1$  的第 15 列安排因素  $F'$ ， $E$  的第 14 列安排因素  $G'$ 。利用第 1 列赋闲列，把  $C$ 、 $D$  配列于第 10、11 列和第 2、3 列，得点线图 46.2-10。

画点线图的顺序如下：

- 1) 第 1 列对应赋闲列。

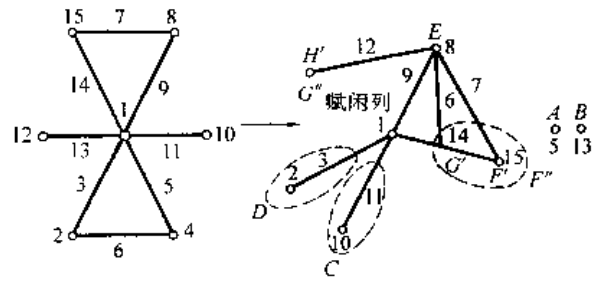


图 46.2.10  $L_{16}$  (2) 型点线图及其变形

- 2) 把连接第 1 列与 15 列的线段对应于  $F''$ 。
- 3) 把  $E$  对应的第 8 列与第 1、15、14 列连接起来，去掉第 9、6、7 列。
- 4) 把  $F'$ 、 $G'$  对应于  $E_1$  的第 15、14 列。
- 5) 把第 8 列与第 4 列连接起来，去掉它们的交互列即第 12 列，使  $H'$  对应于第 8 列为第一水平的第 4 列，把  $G''$  对应于第 8 列的第二水平的第 4 列。
- 6) 把  $C$  对应于赋闲列与第 10 列的连线，把  $D$  对应于赋闲列与第 2 列的连线。

如还有其他因素，可以把它们与赋闲列进行混杂进行配列。

(3) 把因素配列于正交表

把图 46.2-10 的点线图与正交表  $L_{16}$  对应，得到表 46.2-22。

表 46.2-22 拟因素法的配列

因素列 No	赋闲列		D		H' G''		A	6	7	E	赋闲列		C		B	G'	F' F''	数据
	1	2	3	4	5	8	9	10	11	12	13	14	15					
1	1	1		$H'_1$	1		1	1	1		1	$G'_1$	$F'_1$					7
2	1	1		$G''_1$	1		2	2	2		2	—	$F''_2$					-4
3	1	1		$H'_2$	2		1	1	1		2	$G'_2$	$F'_2$					-14
4	1	1		$G''_2$	2		2	2	2		1	—	$F''_1$					3
5	1	2		$H'_1$	1		1	1	2		1	$G'_2$	$F'_2$					-8
6	1	2		$G''_1$	1		2	2	1		2	—	$F''_1$					11
7	1	2		$H'_2$	2		1	1	2		2	$G'_1$	$F'_1$					-13
8	1	2		$G''_2$	2		2	2	1		1	—	$F''_2$					-3
9	2	2		$H'_1$	2		1	2	2		2	$G'_1$	$F'_2$					11
10	2	2		$G''_1$	2		2	1	3		1	—	$F''_2$					-8
11	2	2		$H'_2$	1		1	2	2		1	$G'_2$	$F'_1$					-5
12	2	2		$G''_2$	1		2	1	3		2	—	$F''_2$					-2
13	2	3		$H'_1$	2		1	2	3		2	$G'_2$	$F'_1$					11
14	2	3		$G''_1$	2		2	1	2		1	—	$F''_3$					-4
15	2	3		$H'_2$	1		1	2	3		1	$G'_1$	$F'_2$					-15
16	2	3		$G''_2$	1		2	1	2		2	—	$F''_2$					-3

2.4.4 方差分析

实验测量得到的灵敏度数据如表 46.2-23 数据栏所示。单位是 dB (分贝)，已减去虚拟平均值 90 后的数据。

首先求出各因素不同水平的测量值之和。计算波

动如下：

$$CF = \frac{(-36)^2}{16} = 81.0 \quad (f=1)$$

$$S_A = \frac{(19 \cdot 17)^2}{16} = 0.2 \quad (f=1)$$

$$S_B = \frac{(-33+3)^2}{16} = 56.2 \quad (f=1)$$

表 46.2-23 不同水平测量值之和

	A	B	E	F'	G'	H'	G''	第 1 列	第 9 列
第 1 水平	-19	-33	-26	0	-10	21	-5	-21	-45
第 2 水平	-17	-3	-10	-26	-16	-47	-5	-15	9
C 总计			D 总计			F'' 总计			
(赋闲列) <sub>1</sub> C <sub>1</sub> (1 C <sub>2</sub> (-22			(赋闲列) <sub>1</sub> (D <sub>1</sub> -8 D <sub>2</sub> -13			(赋闲列) <sub>1</sub> (F'' <sub>1</sub> 14 F'' <sub>2</sub> -7			
(赋闲列) <sub>2</sub> C <sub>2</sub> (-1 C <sub>3</sub> (-14			(赋闲列) <sub>2</sub> (D <sub>2</sub> -4 D <sub>3</sub> -11			(赋闲列) <sub>2</sub> (F'' <sub>2</sub> -11 F'' <sub>3</sub> -6			

$$S_{C_1 C_2} = \frac{(1+22)^2}{8} = 66.1 \quad (f=1)$$

$$S_{C_2 C_3} = \frac{(-1+14)^2}{8} = 21.1 \quad (f=1)$$

$$S_{D_1 D_2} = \frac{(-8+13)^2}{8} = 3.1 \quad (f=1)$$

$$S_{D_2 D_3} = \frac{(-4+11)^2}{8} = 6.1 \quad (f=1)$$

$$S_E = \frac{(-26+10)^2}{16} = 16.0 \quad (f=1)$$

$$S_{F'} = \frac{(0+26)^2}{8} = 84.5 \quad (f=1)$$

$$S_{G'} = \frac{(-10+16)^2}{8} = 4.5 \quad (f=1)$$

$$S_{H'} = \frac{(21+47)^2}{8} = 578.0 \quad (f=1)$$

$$S_{F''_1 F''_2} = \frac{(14+7)^2}{4} = 110.2 \quad (f=1)$$

$$S_{F''_2 F''_3} = \frac{(-11+6)^2}{4} = 6.2 \quad (f=1)$$

$$S_{G''} = \frac{(-5+5)^2}{8} = 0.0 \quad (f=1)$$

$$S_{1列} = \frac{(-21+15)^2}{16} = 2.2 \quad (f=1)$$

$$S_{2列} = \frac{(-45-9)^2}{16} = 182.2 \quad (f=1)$$

上述各因素的波动之和为 1136.6, 总波动为:

$$S_T = 7^2 + (-4)^2 + \dots + (-3)^2 - 81.0 = 1137.0 \quad (f=15)$$

把较小因素 A、D<sub>1</sub>、D<sub>2</sub>、D<sub>2</sub>、D<sub>3</sub>、G'、F''<sub>1</sub>F''<sub>2</sub>、G''、赋闲列

的第一列归并起来, 求自由度为 7 的误差方差:

$$S_{(e)} = 2.2 + 0.2 + 3.1 + 6.1 + 4.5 + 6.2 + 0.0 = 22.3 \quad (f=7)$$

得方差分析表 46.2-24。

2.4.5 因素显著性估计与最宜条件

(1) 估计显著性

得

$$\begin{cases} B_1 = 90 + \frac{-33}{8} = 85.9 \pm 1.5 \\ B_2 = 90 + \frac{-3}{8} = 89.6 \pm 1.5 \end{cases}$$

表 46.2-24 碳粒的方差分析表

要 因	f	S	V	S'	p(%)
赋闲 1 列	1	2.2	2.2○		
赋闲 9 列	1	182.2	182.2	179.0	15.7
A	1	0.2	0.2○		
B	1	56.2	56.2	53.0	4.7
C (C <sub>1</sub> C <sub>2</sub> )	1	66.1	66.1	62.9	5.5
	1	21.1	21.1	17.9	1.6
D (D <sub>1</sub> D <sub>2</sub> )	1	3.1	3.1○		
	1	6.1	6.1○		
	1	16.0	16.0△	12.8	1.1
E	1	16.0	16.0	81.3	7.2
F'	1	84.5	84.5		
G'	1	4.5	4.5○		
H'	1	578.0	578.0	574.8	50.6
F'' (F'' <sub>1</sub> F'' <sub>2</sub> )	1	110.2	110.2	107.0	9.4
	1	6.2	6.2○		
G''	1	0.0	0.0○		
e	0	—			
打有“○”的(e)	(7)	(22.3)	(3.2)	47.9	4.2
总计	15	1136.6		1136.6	100.0

$$\begin{cases} C_1 = 90 + \frac{1}{4} = 90.2 \rightarrow 92.85 \pm 2.1 \\ C_2 = 90 + \frac{-22}{4} = 84.5 \end{cases} \left. \begin{array}{l} \\ \\ \end{array} \right\} 87.15 \pm 2.1$$

$$\begin{cases} C_2 = 90 + \frac{-1}{4} = 89.8 \\ C_3 = 90 + \frac{-14}{4} = 86.5 \rightarrow 83.85 \pm 2.1 \end{cases}$$

(46.2-3)

$$\begin{cases} E_1 = 90 + \frac{-26}{8} = 86.8 \pm 1.5 \\ E_2 = 90 + \frac{-10}{8} = 88.8 \pm 1.5 \end{cases}$$

$$\begin{cases} F'_1 = 90 + \frac{0}{4} = 90.0 \pm 2.1 \\ F'_2 = 90 + \frac{-26}{4} = 83.5 \pm 2.1 \end{cases}$$

$$\begin{cases} H'_1 = 90 + \frac{21}{4} = 95.2 \pm 2.1 \\ H'_2 = 90 + \frac{-47}{4} = 78.2 \pm 2.1 \end{cases}$$

$$\begin{aligned}
 & \left. \begin{aligned}
 & \text{(赋闲列)}_1 \left\{ \begin{aligned}
 & F_1'' = 90 + \frac{14}{2} = 97.0 \rightarrow 96.0 \pm 3.0 \\
 & F_2'' = 90 + \frac{-7}{2} = 86.5 \\
 & F_2'' = 90 + \frac{-11}{2} = 84.5 \\
 & F_3'' = 90 + \frac{-6}{2} = 87.0 \rightarrow 88.0 \pm 3.0
 \end{aligned} \right\} 85.5 \pm 3.0 \\
 & \text{(赋闲列)}_2
 \end{aligned} \right\} \\
 & \text{(46.2-4)}
 \end{aligned}$$

$$\begin{aligned}
 & \left. \begin{aligned}
 & (\bar{9})_1 = 90 + \frac{-45}{8} = 84.4 \pm 1.5 \\
 & (\bar{9})_2 = 90 + \frac{9}{8} = 91.1 \pm 1.5
 \end{aligned} \right\} \\
 & \text{(46.2-5)}
 \end{aligned}$$

修正后的因素 C 的主效应, 只有 C<sub>1</sub>、C<sub>2</sub> 的比较与 C<sub>2</sub>、C<sub>3</sub> 的比较是正交表数据。因此实际的反复数为 4。

$$\pm \sqrt{\frac{5.59+3.2}{4}} \approx \pm 1.5$$

将计算结果画成曲线图。(见图 46.2-11)

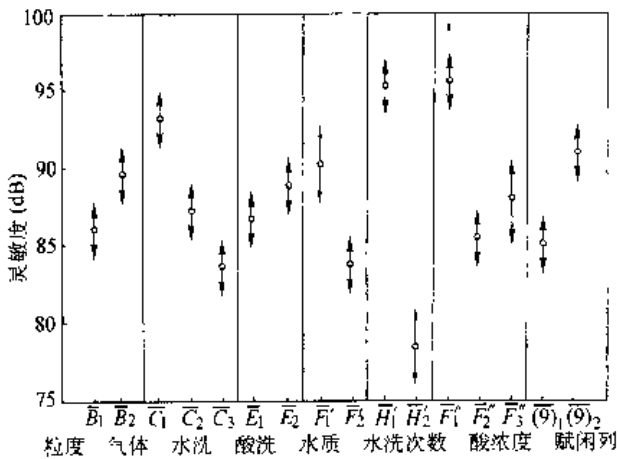


图 46.2-11 因素效应图

(2) 求效率最大的条件

首先比较水洗 E<sub>1</sub>, 酸洗 E<sub>2</sub> 的最宜条件。

(1) E<sub>1</sub> 的最宜条件 E<sub>1</sub>F<sub>1</sub>H<sub>1</sub> 的  $\hat{\mu}$  为

$$\begin{aligned}
 \hat{\mu} &= F_1' + H_1' - E_1 - 90.0 + 95.2 - 86.8 \\
 &= 98.4 \pm 2.6
 \end{aligned}$$

(2) E<sub>2</sub> 的最宜条件 E<sub>2</sub>F<sub>1</sub> 的  $\hat{\mu}$  为

$$\hat{\mu} = E_2 + (F_1' - E_2) = F_1' = 96.0 \pm 3.0$$

水洗的最宜条件并不显著, 但比酸洗的最宜条件好。而且酸洗成本高, 因此最宜条件为 E<sub>1</sub>F<sub>1</sub>H<sub>1</sub>。整个实验的最宜条件为 B<sub>2</sub>C<sub>1</sub>E<sub>1</sub>F<sub>1</sub>H<sub>1</sub>, 工序平均  $\hat{\mu}$  为

$$\begin{aligned}
 \hat{\mu} &= \bar{T} + (\bar{B}_2 - \bar{T}) + (\bar{C}_1 - \bar{T}) + (\bar{E}_1 F_1 H_1 - \bar{T}) \\
 &= \bar{B}_2 + \bar{C}_1 + \bar{E}_1 F_1 H_1 - 2\bar{T} \\
 &= 89.6 + 92.8 + 98.4 - 2 \times 87.8 \\
 &= 105.2
 \end{aligned}$$

在最宜条件下, 各个 x 值的预测值, 由下式求得:

$$\begin{aligned}
 x &= \hat{\mu} \pm \sqrt{F \times V_e \times \left( \frac{1}{n_e} + 1 \right)} \\
 &= 105.2 \pm \sqrt{5.59 \times 3.2 \left( \frac{9}{16} + 1 \right)} \\
 &= 105.2 \pm 5.3
 \end{aligned}$$

有效反复数如下:

$$\begin{aligned}
 \frac{1}{n_e} &= \frac{1(m) + 1(B) + 2(C) + 5(E, F, H)}{19} \\
 &= \frac{9}{16}
 \end{aligned}$$

## 2.5 粉碎过程实验

### 2.5.1 配列与数据

考虑多水平因素的问题。例如, 可塑剂有 13 种, 或者结晶反应中的添加剂有 17 种等等, 出现多水平因素的情况, 像种类那样与质量有关的因素要比数量有关的因素多。然而有各种多水平因素时, 怎样进行实验设计, 拟因素法的赋闲列法是有效的。下例解释赋闲列法在粉碎过程实验中的应用, 粉碎工序中的七水平因素 A, 三水平因素 B、C、D, 二水平因素 E、F 的主效应, 在 L<sub>16</sub> 中的配列。作为四水平拟因素 A', 对应于适当的赋闲列中去, 作表 46.2-25 的配列。

表 46.2-25

拟因素 赋闲列	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
1	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
2	A <sub>1</sub>	A <sub>5</sub>	A <sub>6</sub>	A <sub>7</sub>

因素排列后可以从赋闲列的一水平和二水平实验分别对 A<sub>1</sub>、A<sub>2</sub>、A<sub>3</sub>、A<sub>4</sub> 和 A<sub>1</sub>、A<sub>5</sub>、A<sub>6</sub>、A<sub>7</sub> 进行比较, 那么就可以通过共同的水平 A<sub>1</sub> 对全部水平进行比较。这时称 A<sub>1</sub> 为对照水平, 但对照水平未必一定是 A<sub>1</sub>, 通常可选最重要的水平; 当因素 A 有六个水平时, 按照表 46.2-26, 对照水平可取两个。

表 46.2-26

拟因素 赋闲列	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
1	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
2	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>

这种方法与系统分组设计或对照法十分相似, 但采用拟因素法进行数据分析时, 只需利用正交数据, 这点是区别所在。通过最小二乘法利用全部数据的方法, 计算复杂且精度提高不多, 比起不完全区组设计法, 拟因素法更为方便。即使采用了拟因素法, 还可以采用最小二乘法把部分混杂的赋闲列的数据信息分离, 但是通常这种数据只占正交数据的几分之一, 因而复杂的计算能分离出一些数据, 但是精度并没有提高。当然,

对不存在正交数据的数据来说,即使能分离少量的情报也是有意义的。因而在采用拟因素法后往往也做这种处理。但是当存在正交数据时,增加几分之一的数据信息,即使置信界限缩小,也没有多大意义。

### 2.5.2 数据分析

按表 46.2-27 的因素配列,进行 16 种实验。从单位时间的产量(kg/min)中减去虚拟平均值 20 所得的数据列于表的数据栏中。

表 46.2-27 正交表  $L^0$  的  $7 \times 3^3 \times 2^2$  型的因素配列

因素列 No	赋闲列	A			B		C		D		E		F	数据
		2	3	4	8	9	10	11	12	13	14	15		
1	1	1	1	1	1	1	1	1	1	1	1	1	2	
2	1	1	2	2	2	2	2	2	2	2	2	2	6	
3	1	2	1	1	2	2	2	2	2	2	2	2	1	
4	1	2	2	2	2	1	1	1	1	1	1	1	2	
5	1	3	1	1	1	1	2	2	2	2	2	2	4	
6	1	3	2	2	2	2	1	1	1	1	1	1	7	
7	1	4	1	1	2	2	1	1	1	1	1	1	6	
8	1	4	2	2	2	1	2	2	2	2	2	2	1	
9	2	1	1	1	1	1	1	1	2	2	2	2	1	
10	2	1	3	3	3	3	2	2	2	2	2	2	8	
11	2	5	1	1	1	3	2	2	1	1	1	1	4	
12	2	5	3	3	3	1	1	1	2	2	2	2	1	
13	2	6	1	3	3	1	2	2	1	1	1	1	0	
14	2	6	3	1	3	1	2	2	1	1	1	1	7	
15	2	7	1	3	3	1	2	2	1	1	1	1	2	
16	2	7	3	1	3	1	2	2	1	1	1	1	0	
总计													52	

		辅助表							
(赋闲列)	A	8	A <sub>1</sub>	B <sub>1</sub>	C <sub>1</sub>	D <sub>1</sub>	E <sub>1</sub>	F <sub>1</sub>	
1	(赋闲列)	A <sub>2</sub>	3	A <sub>3</sub>	B <sub>2</sub>	C <sub>2</sub>	D <sub>2</sub>	E <sub>2</sub>	F <sub>2</sub>
2	(赋闲列)	A <sub>3</sub>	11	A <sub>6</sub>	B <sub>1</sub>	C <sub>1</sub>	D <sub>1</sub>		
总计	52	A <sub>4</sub>	7	A <sub>7</sub>	B <sub>3</sub>	C <sub>2</sub>	D <sub>3</sub>		
			2		16	11	21		

$$CF = \frac{52^2}{16} = 169.0$$

$$S_D = \frac{(29-23)^2}{16} = 2.2$$

$$S_{A_1-A_4} = \frac{8^2 + 3^2 + 11^2 + 7^2}{2} - \frac{29^2}{8} = 16.4$$

$$S_{A_1A_3A_6A_7} = \frac{9^2 + 3^2 - 7^2 + 2^2}{2} - \frac{23^2}{8} = 13.4$$

$$S_{B_1B_3} = \frac{(13-16)^2}{8} = 1.1$$

$$S_{B_1B_3} = \frac{(7-16)^2}{8} = 10.1$$

$$S_{C_1C_2} = \frac{(11-18)^2}{8} = 6.1$$

$$S_{C_1C_2} = \frac{(12-11)^2}{8} = 0.1$$

$$S_{D_1D_2} = \frac{(9-20)^2}{8} = 15.1$$

$$S_{D_1D_3} = \frac{(2-21)^2}{8} = 45.1$$

$$S_E = \frac{(28-24)^2}{16} = 1.0$$

$$S_F = \frac{(29-23)^2}{16} = 2.3$$

由此得方差分析表(见表 46.2-28)。

表 46.2-28 方差分析表

要因	f	S	V
赋闲列	1	2.2	2.2
A (A <sub>1</sub> ~A <sub>3</sub> )	3	16.4	5.5
A (A <sub>1</sub> A <sub>3</sub> A <sub>6</sub> A <sub>7</sub> )	3	13.4	4.5
B (B <sub>1</sub> B <sub>2</sub> )	1	1.1	1.1
B (B <sub>1</sub> B <sub>3</sub> )	1	10.1	10.1
C (C <sub>1</sub> C <sub>2</sub> )	1	6.1	6.1
C (C <sub>1</sub> C <sub>2</sub> )	1	0.1	0.1
D (D <sub>1</sub> D <sub>2</sub> )	1	15.1	15.1
D (D <sub>1</sub> D <sub>3</sub> )	1	45.1	45.1
E	1	1.0	1.0
F	1	2.3	2.3
e	0	0.0	0.0
标有“○”的(e)	(8)	(20.0)	(2.50)
总计	15	112.9	

这时如用带有“○”的自由度占一半左右的 V 值较小的因素并起来,对其他因素进行检验,只有 0 有显著性。由于正交表中只有 D 显著,所以最好改变因素配列,把 D<sub>1</sub>D<sub>2</sub>D<sub>3</sub> 作为一个因素,并整理方差分析表。

$$S_D = \frac{11^2}{8} + \frac{20^2}{4} + \frac{21^2}{4} - CF = 56.4 \quad (f=2)$$

$$S_e = S_T - S_D = 5.6 \quad (f=13)$$

$$\bar{D}_1 = 20 + \frac{11}{8} = 21.4 \pm 1.3$$

$$\bar{D}_2 = 20 - \frac{20}{4} = 25.0 \pm 1.8$$

$$\bar{D}_3 = 20 + \frac{21}{4} = 25.2 \pm 1.8$$

### 2.6 三轮摩托轮胎寿命实验

在三轮车轮胎寿命实验中,当有两个七水平因素 A、B,三水平因素 C,两个二水平因素 D、E 时,只考虑求主效应的实验。假定实验采用九辆摩托车进行实验,不附加任何其他限制条件,可以在任意道路上行使。假设实验时间为三个月,则在三个月时间里让车随便行使的实验条件会随车的不同发生很大的变化。

现设区组为 R<sub>1</sub>, R<sub>2</sub>, ..., R<sub>9</sub>, 轮胎的位置为 V<sub>1</sub> (前), V<sub>2</sub> (后左), V<sub>3</sub> (后右)。只要同一辆车,可以认



为  $V_2, V_3$  的条件相同, 但是由于车不是在道路正中行驶, 一般便存在车的一侧负荷大的倾向, 因此  $R$  是区组大小为 3 的区组因素,  $V$  是三水平的标示因素。

如果按正交表  $L_{27}$  进行因素配列, 那么由于  $R$  的区组大小为 3, 所以  $R$  是九水平的, 因此我们的问题就是除  $7^2 \times 3^2 \times 2^2$  的控制因素  $A, B, C, D, E$  之外, 必须把九水平的  $R$  和三水平  $V$  配列于正交表  $L_{27}$  中。

可这时  $R$  的主效应没有技术上的意义, 所以对于  $R$  只能在一定程度上对区组间的方差作出估计就可以了,  $R$  的一部分效应混杂着控制因素的效应, 是没有丝毫关系的。当然反过来就必须避免在控制因素中混杂有区组因素  $R$  的效应。

区组因素的特点就是不能构成对照水平和组合因素, 因此因素配列如图 46.2-12 所示。

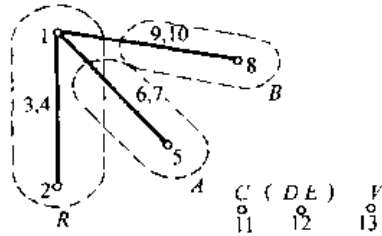


图 46.2-12 因素配列型

表 46.2-29

拟因素	$A_1^i$	$A_2^i$	$A_3^i$
1	$A_1$	$A_2$	$A_3$
2	$A_1$	$A_4$	$A_5$
3	$A_1$	$A_6$	$A_7$

a)

拟因素	$B_1^i$	$B_2^i$	$B_3^i$
1	$B_1$	$B_2$	$B_3$
2	$B_1$	$B_4$	$B_5$
3	$B_1$	$B_6$	$B_7$

b)

其中,  $D, E$  是组合因素,  $A, B$  是拟因素, 它们组成如表 46.2-29 所示配列:

第一列是赋闲列 ( $\tau$ ) 是把其九水平代入第 1 列与第 2 列的九种组合中的因素。因此要估计  $\sigma_R^2$ , 自由度只能用 6, 因素配列如表 46.2-30 所示。不仅七水平的, 而且六、五、四等水平的情形, 可以同时并用拟水平和组合法。

现在 27 种不同条件下各生产两个轮胎, 用 18 台三轮摩托车, 测量刚制成的轮胎的不平衡量为  $k_0$ , 修正后的不平衡量为  $k_1$ , 一个月后的不平衡量为  $k_2$ , 两个月后的不平衡量为  $k_3$ 。对于  $R_{10}, R_{11}, \dots, R_{18}$  做了与  $R_1, R_2, \dots, R_9$  同样的实验。

表 46.2-30 大小为 3 的赋闲因素和三水平的标示因素  $7^2 \times 3 \times 2^2$  型的因素配列

No.	R			A			B			C	(DE)	V	$K_0$	$K_1$	$K_2$	$K_3$
	1	2	3	4	5	6	7	8	9	10	11	12				
1	1	10	1	1	1	1	1	11	1	11	1	=	=	=	=	
2	1	10	2	2	2	2	2	21	2	21	2	⋮	⋮	⋮	⋮	
3	1	10	3	3	3	3	3	31	3	31	3	⋮	⋮	⋮	⋮	
4	2	11	1	2	2	2	3	12	3	12	3	⋮	⋮	⋮	⋮	
5	2	11	2	3	1	1	1	11	1	11	1	⋮	⋮	⋮	⋮	
6	2	11	3	1	2	2	2	21	2	21	2	⋮	⋮	⋮	⋮	
7	3	12	1	3	2	2	1	21	2	21	2	⋮	⋮	⋮	⋮	
8	3	12	2	1	3	1	3	12	3	12	3	⋮	⋮	⋮	⋮	
9	3	12	3	2	2	1	1	11	1	11	1	⋮	⋮	⋮	⋮	
10	4	13	1	1	1	1	1	11	1	11	1	⋮	⋮	⋮	⋮	
11	4	13	4	4	4	4	4	21	4	21	4	⋮	⋮	⋮	⋮	
12	4	13	5	5	5	5	5	31	5	31	5	⋮	⋮	⋮	⋮	
13	5	14	1	4	4	4	3	11	2	11	2	⋮	⋮	⋮	⋮	
14	5	14	4	3	1	1	1	21	3	21	3	⋮	⋮	⋮	⋮	
15	5	14	5	1	2	2	2	12	1	12	1	⋮	⋮	⋮	⋮	
16	6	15	1	5	2	2	2	12	1	12	1	⋮	⋮	⋮	⋮	
17	6	15	4	1	3	1	3	11	2	11	2	⋮	⋮	⋮	⋮	
18	6	15	5	4	1	1	1	21	3	21	3	⋮	⋮	⋮	⋮	
19	7	16	1	1	1	1	1	11	2	11	2	⋮	⋮	⋮	⋮	
20	7	16	6	6	6	6	2	11	3	11	3	⋮	⋮	⋮	⋮	
21	7	16	7	7	7	7	3	21	1	21	1	⋮	⋮	⋮	⋮	
22	8	17	1	6	3	3	3	21	1	21	1	⋮	⋮	⋮	⋮	
23	8	17	6	7	7	7	1	12	2	12	2	⋮	⋮	⋮	⋮	
24	8	17	7	1	2	2	2	11	3	11	3	⋮	⋮	⋮	⋮	
25	9	18	1	7	2	2	2	11	3	11	3	⋮	⋮	⋮	⋮	
26	9	18	6	1	3	3	3	21	1	21	1	⋮	⋮	⋮	⋮	
27	9	18	7	6	1	1	1	12	2	12	2	⋮	⋮	⋮	⋮	

数据分析方法

数据  $K_0$  与  $K_1, K_2, K_3$  分别进行计算。这里只列出  $K_1, K_2, K_3$  的寿命实验计算式。

(1)  $i$  是第一列的  $i$  水平的合计值。

$$S_m = \frac{(\text{合计})^2}{162} \quad (f=1)$$

$$S_c = S_{1,11} = \frac{(1)_1^2 + (1)_2^2 + (1)_3^2}{54} - S_m \quad (f=2)$$

$$S_R = \frac{R_1^2 + R_2^2 + R_3^2 + \dots + R_{18}^2}{9} - S_m - S_c \quad (f=15)$$

$$S_{A_1 A_2 A_3} = \frac{A_1^2 + A_2^2 + A_3^2}{18} - \frac{(1)_3^2}{54} \quad (f=2)$$

$$S_{A_1 A_4 A_5} = \frac{A_1^2 + A_4^2 + A_5^2}{18} - \frac{(1)_3^2}{54} \quad (f=2)$$

$$S_{A_1 A_6 A_7} = \frac{A_1^2 + A_6^2 + A_7^2}{18} - \frac{(1)_3^2}{54} \quad (f=2)$$

同样计算  $S_B$ , 得

$$S_C = \frac{C_1^2 + C_2^2 + C_3^2}{54} - S_m \quad (f=2)$$

$$S_{(DE)} = \frac{(DE)_1^2 + (DE)_2^2 + (DE)_3^2}{54} - S_m \quad (f=2)$$

$$S_D = \frac{(D_1 E_1 - D_2 E_1)^2}{108} \quad (f=2)$$

$$S_{E^2} = \frac{(D_1 E_1 - D_2 E_2)^2}{108} \quad (f=2)$$

同样计算  $S_V$  与  $S_C$ , 得

$$S_A = \frac{K_1^2 + K_2^2 + K_3^2}{54} S_m \quad (f=2)$$

$$S_{\tau XK} = \frac{\tau_1 K_1^2 + \tau_2 K_2^2 + \dots + \tau_r K_r^2}{18} - S_m - S_r - S_K \quad (f=4)$$

$$S_{RXX} = \frac{R_1 K_1^2 + R_2 K_2^2 + \dots + R_{18} K_{18}^2}{3} - S_m - S_K - S_{\tau XK} - S_r - S_R \quad (f=12)$$

$$S_{A_1 A_2 A_3 XK} = \frac{A_1 K_1^2 + \dots + A_3 K_3^2}{18} - \frac{(1)^2}{54} - S_{A_1 A_2 A_3} - S'_K \quad (f=4)$$

其他因素的求法相同。\$S\_K\$ 是赋闲列为第 1 列的 \$k\$ 的主效应。

实验间的波动为：

$$S_{T_1} = \frac{1}{6} [(第 1 号实验的总计)^2 + \dots + (第 27 号实验的总计)^2] - S_m \quad (f=26)$$

轮胎间的波动为：

$$S_{T_2} = \frac{1}{3} [54 个轮胎每个合计值的平方和] - S_m$$

总波动为：

$$S_T = 162 个数据之间的平方和$$

因此方差分析表中自由度的分配如表 46.2-31 所示。

表 46.2-31 方差分析表

要 因		f		
轮胎间 \$T_2\$	试验间 \$T_1\$	\$m\$	1	
		赋闲列	2	
		\$R\$	15	
		\$V\$	2	
		\$A\$	\$A_1, A_2, A_3\$	2
			\$A_4, A_5, A_6\$	2
			\$A_7, A_8, A_9\$	2
		\$B\$	\$B_1, B_2, B_3\$	2
			\$B_4, B_5, B_6\$	2
			\$B_7, B_8, B_9\$	2
		\$C\$	2	
		\$(DE)\$	2	
		\$e_1\$	0	
		\$e_2\$	18	
		\$K\$	2	
		赋闲列(\$\tau\$)	4	
		\$R \times K\$	30	
		\$V \times K\$	4	
		\$A \times K\$	\$A_1, A_2, A_3 \times K\$	4
			\$A_4, A_5, A_6 \times K\$	4
\$A_7, A_8, A_9 \times K\$	4			
\$B \times K\$	\$B_1, B_2, B_3 \times K\$	4		
	\$B_4, B_5, B_6 \times K\$	4		
	\$B_7, B_8, B_9 \times K\$	4		
\$C \times K\$	4			
\$(DE) \times K\$	4			
\$(e_3)\$	36			
计	162			

### 3 部分追加设计

在一个因素有很多水平且难以配列时,可用部分追加法。

#### 3.1 方差预期值的计算法则

设用正交表 \$L\_N\$ 进行部分追加, \$M < N\$, 在 \$(M + N)\$ 种条件下进行实验。这时除部分追加所用因素外, 方差期望值计算中的因素, 都与使正交表 \$L\_N\$ 完全反复的情况相同。部分追加因素只要分为反复计算部分与不完全反复计算部分就行了。实验间的误差作为误差过剩部分, 可把 \$(N - M)\$ 份误差按正交表 \$L\_N\$ 的第一组的各自由度平均地进行分配。当有重复误差时, 可根据重复试验的实际情况计算重复误差。例如, 设 \$A\$ 为五水平因素, \$B, C, D, E\$ 均为二水平因素。当用因素 \$A\$ 做正交表 \$L\_8\$ 的部分追加时, 方差期望按下述方法计算。因为是在 \$L\_8\$ 中做 \$M = 2\$ 的部分追加, 所以误差过剩部分等于 \$(8 - 2)\sigma^2 = 6\sigma^2\$。这个 \$\sigma^2\$, 可按 \$L\_8\$ 的因素 \$m, A, B, C, D, E\$ 的自由度 1, 3, 1, 1, 1 平均分配, 即每个自由平均分配 \$6\sigma^2/8\$。因而进入 \$S\_m, S\_A, S\_B, \dots, S\_E\$ 的误差方差等于: \$(1 + 6/8)\sigma^2 = (7/4)\sigma^2, (4 + 3 \times 6/8)\sigma^2 = (25/4)\sigma^2, (1 + 6/8)\sigma^2 = (7/4)\sigma^2, \dots, (1 + 6/8)\sigma^2 = (7/4)\sigma^2\$。另一方面由于因素 \$m, B, C, D, E\$ 与作两次实验的情况相同, 因而有 \$16m^2, 8\sigma\_B^2, 8\sigma\_C^2, 8\sigma\_D^2, \sigma\_E^2\$。对于因素 \$A\$ 来说, 只有 \$A\_1, A\_2, A\_3\$ 考虑到两倍反复, 作两次试验, \$A\$ 可考虑进行两次试验时的不完全反复。根据 4, 4, 4, 2, 2 的调和平均等于 \$20/7\$, 所以得到 \$E(S) = 25/4\sigma^2 + (A \text{ 的自由度}) \times 20/7\sigma^2\$。

#### 3.2 实验例子

设 \$A\$ 为五水平因素, \$B, C, D, E\$ 均为二水平因素。现在只考虑主效应, 把部分追加法用于 \$L\_8\$ 的正交表。对合计 10 种条件下的所产物品的粘度, 各进行三次重复测量, 所得测量值(减去拟平均值 50 后的数据)如表 46.2-32 所示。试求解下述问题:

表 46.2-32 部分追加法的配列与数据

序号	A			B	C	D	E	数据			小计
	1	2	3					(-50)			
1	1	1	1	1	1	1	6	4	5	15	
2	1	2	2	2	2	2	2	0	1	3	
3	2	1	1	2	2	9	8	11	28		
4	2	2	2	1	1	1	0	-1	0		
5	3	1	2	1	2	14	13	14	41		
6	3	2	1	2	1	7	4	5	16		
7	4	1	2	2	1	9	8	12	29		
8	4	2	1	1	2	6	6	4	16		

(续)

序号	A			B	C	D	E	数据			小计	计
	1	2	3					(-50)				
1	1	1	1	1	1	1	6	4	5	15	30	
2	1	2	2	2	2	2	2	0	1	3	6	
3	2	1	1	2	2	2	9	8	11	28	56	
4	2	2	2	1	1	1	1	0	-1	0	0	
5	3	1	2	1	2	2	14	13	14	41	82	
6	3	2	1	2	1	1	7	4	5	16	32	
7	5	1	2	2	1	1	10	13	11	34	63	
8	5	2	1	1	2	2	9	10	7	26	42	

- 1) 方差分析;
- 2) 估计主要因素的效应;
- 3) 在使粘度达到最大的组合条件下,所有物品的测量值的存在范围。

3.2.1 方差分析

$$CF = \frac{(30+6+\dots+42)^2}{48} = 2015.02 \quad (f=1)$$

$$S_{T_1} = \frac{1}{3} [2(15^2+3^2+\dots+16^2)+29^2+16^2+34^2+26^2] - CF = 931.31 \quad (f=9)$$

$$S_A = \frac{36^2+56^2+114^2+45^2+60^2}{12} + \frac{45^2+60^2}{6} - CF = 374.81 \quad (f=4)$$

式中  $A_1 = 30+6 = 36, A_2 = 56+0 = 56, \dots, A_5 = 34+$

26+60

$$S_B = \frac{(231-80)^2}{48} = 475.02 \quad (f=1)$$

式中  $B_1 = 30+56+82+63 = 231, B_2 = 6+0+32-42-80$

$$S_C = \frac{(160-151)^2}{48} = 1.69 \quad (f=1)$$

$$S_D = \frac{(154-157)^2}{48} = 0.19 \quad (f=1)$$

$$S_E = \frac{(125-186)^2}{48} = 77.52 \quad (f=1)$$

$$S_{T_1} = S_{T_1} (S_A + S_B - S_C + S_D + S_E) - 2.08 \quad (f=1)$$

$$S_{T_2} = \left[ 6^2 - 4^2 - 5^2 - \frac{15^2}{3} \right] + \left[ 2^2 + 0^2 + 1^2 - \frac{3^2}{3} \right] - \dots + \left[ 9^2 + 10^2 + 7^2 - \frac{26^2}{3} \right] = 36.67 \quad (f=20)$$

No. 1~No. 6 不做重复误差的双重计算。因为重复处于正交状态,双重计算不必要。

$$S_T = S_{T_1} + S_{T_2} = 931.31 + 36.67 = 967.98$$

有上列计算值,得方差分析表(见表 46.2-33)。(e) 的波动并项,用修正后的方差进行计算,即

$$S_{(e)} = 0.96 + 0.11 + 2.08 + 36.67 = 39.82 \quad (f=23)$$

表 46.2-33 方差分析表

因素	f	S	V	E(V)	K (修正系数)	KV
A	4	374.81	93.70	$\frac{25}{16}(\sigma_1^2 + 3\sigma_1^2) - \frac{60}{7}\sigma_A^2$	$\frac{16}{25}$	59.97
B	1	475.02	475.02	$\frac{7}{4}(\sigma_2^2 + 3\sigma_2^2) + 24\sigma_B^2$	$\frac{4}{7}$	271.44
C	1	1.69	1.69	$\frac{7}{4}(\sigma_3^2 + 3\sigma_3^2) - 24\sigma_C^2$	$\frac{4}{7}$	0.96
D	1	0.19	0.19	$\frac{7}{4}(\sigma_4^2 + 3\sigma_4^2) + 24\sigma_D^2$	$\frac{4}{7}$	0.11
E	1	77.52	77.52	$\frac{7}{4}(\sigma_5^2 + 3\sigma_5^2) - 24\sigma_E^2$	$\frac{4}{7}$	44.30
e <sub>1</sub>	1	2.08	2.08	$\sigma_2^2 + 3\sigma_1^2$	1	2.08
e <sub>2</sub>	20	36.67	1.84	$\sigma_2^2$	1	1.84
(e)	(23)				(39.82)	(1.73)
计	29	969.98				

3.2.2 显著因素估计

$$\bar{A}_1 = 50 + \frac{18}{6} = 53 \pm 1.1 \left( = \sqrt{\frac{4.28 \times 1.73}{6}} \right)$$

$$\bar{A}_2 = 50 + \frac{28}{6} = 54.7 \pm 1.1$$

$$\bar{A}_3 = 59.5 \pm 1.1$$

$$\bar{A}_4 = 57.5 \pm 1.1$$

$$\bar{A}_5 = 60.0 \pm 1.1$$

在置信界限计算中,由于  $\bar{A}_1, \dots, \bar{A}_5$  都是 6 个数据的平均,所以  $n_e = 6$ 。

4 相互影响部分忽略设计

当几个因素可以连续改变水平时,往往希望把这些因素间及这些因素与其他因素间的交互作用,安排

在第次项截尾。

### 4.1 两水平系统的相互影响部分忽略设计

二水平因素间的交互作用,只有一个自由度,显然不能讨论部分省略。在此,仅限于讨论水平与二水平间的交互作用,四水平间的交互作用。对于四水平因素,都只考虑温度、时间、纯度等连续值的因素,并在原则上取等间隔的四水平。当温度的函数关系是非线形时,从一开始就要用对数表示温度的四水平,如果是等间隔取水平,就称等间隔的四水平。

设因素 A 取四水平, B 取二水平,根据正交多项式的成分,把 A 的效应分解为一次、二次、三次时,相应地记为  $A_1, A_2, A_3$ 。另一方面,不要求因素 B 为连续变量,但在二水平情况下,也可以考虑连续变量,一般也可以取因素 B 为连续变量。至于 A 的主效应,不能忽视一次、二次、三次成分,但就因素 A 与 B 的交互作用来说,只需考虑  $A_1 \times B$ ,而不考虑  $A_2 \times B$  或  $A_3 \times B$ 。

在这种情况下,就无必要求算  $A \times B$  的总计自由度 3,只需求自由度 2,若进行修正的话,则可以只求自由度 1。

例如把因素 A 对应于 L 的 1、2、3 列,把因素对应于 4 列时,  $A \times B$  虽对应于 5、6、7 列,但可以只根据 5 列情报,用下式求算  $A_1 \times B$  的系数。设效率为 80%,  $n_A, n_B$  为因素 A, B 水平间隔。

$$\hat{l}_{11} = \frac{5 \text{ 列 1 之和} - 5 \text{ 列 2 之和}}{8h_A h_B}$$

$$S_{A_1 B} = \frac{(5 \text{ 列 1 之和} - 5 \text{ 列 2 之和})}{16}$$

$$V_{\sigma^2}(\hat{l}_{11}) = \frac{\lambda^2}{4h_A^2 h_B^2}$$

为此作 A 与 B 的二元表,把 5 列的水平记入,如 46.2-34 表所示。

表 46.2-34

	$B_1$	$B_2$
$A_1$	1	2
$A_2$	1	2
$A_3$	2	1
$A_4$	2	1

5 列 1 是  $A_1 B_1, A_2 B_1, A_3 B_2, A_4 B_2$  的和,如果交互作用的成分  $L_{11} (A - \bar{A})(B - \bar{B})$  来考察,则有

$$2[(A_1 - \bar{A})(B_1 - \bar{B}) + (A_2 - \bar{A})(B_2 - \bar{B}) + (A_3 - \bar{A})(B_2 - \bar{B}) + (A_4 - \bar{A})(B_1 - \bar{B})]l_{11}$$

$$= 2 \left[ \left( -\frac{3}{2}h_A \right) \left( -\frac{1}{2}h_B \right) + \left( -\frac{1}{2}h_A \right) \left( \frac{1}{2}h_B \right) + \left( \frac{1}{2}h_A \right) \left( \frac{1}{2}h_B \right) + \left( \frac{3}{2}h_A \right) \left( -\frac{1}{2}h_B \right) \right] l_{11}$$

$$= 4h_A h_B l_{11}$$

同样,5 列 2 的和等于

$$-4h_A h_B l_{11}$$

因而有  $8h_A h_B l_{11} = (1 \text{ 的和}) - (2 \text{ 的和})$

在此情况下,  $A_1 \times B$  的成分的 20% 进入 6 列中,但是  $A_1 \times B$  的成分并未进入 7 列。因为 7 列具有  $A_2 \times B$  的效应。因而在不求  $A_2 \times B, A_3 \times B$  的假定下,7 列可以原样配列其他因素;如把  $A_1 \times B$  的效应进行修正 6 列也可以配列其他因素。

不仅如此,如果同 5 列的  $A_1 \times B$  的效应比较,6 列的  $A \times B$  的效应大小只占 1/4,所以大致以 1% 的风险率判定  $S_{A_1 B}$  的影响时,根据情况可以不作修正。

上述考虑适用于四水平之间的交互作用。为了二水平系完整地提出四水平之间的交互作用,除主效应外,需要 9 条线段,但当只考虑一次与一次的交互作用且不需要修正时,两条线段顶点之间的线段有 4 条就够了,当需要修正时,第 1 列的交互作用有一条就够了。

因而,在 A, B 为连续因素且有四水平, C, D, E, F 有二水平,交互作用只考虑  $A_1 \times B_1, A_1 \times C, B_1 \times C, C \times D, A_1 \times D$  的情况,可按图 46.2-13 的方式配列。

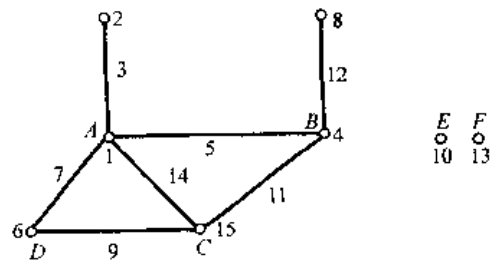


图 46.2-13  $A \times B, A \times C, A \times D, B$

在 A 的水平大小顺序为  $A_1, A_2, A_3, A_4$  时,需要使  $A_1, A_2$  对应于第一列的 1,  $A_3, A_4$  对应于第一列的 2,第一列为头列。B 必须以第四列为头列进行配列,与 A 一样。

这时,如果设  $L_6$  的重复数为 r,则根据 5 列,  $A_1 \times B_1$  的系数  $l_{11}$  的估计值为:

$$\hat{l}_{11} = \frac{5 \text{ 列 1 的和} - 5 \text{ 列 2 的和}}{16r h_A h_B}$$

$A_1 \times B_1$  的系数的波动值为:

$$S_{A_1 \times B_1} = S_{3r}$$

$A_1 \times B_1$  的系数的方差为:

$$V_{\sigma^2}(\hat{l}_{11}) = \frac{\lambda^2}{16r h_A^2 h_B^2}$$

因而求置信界限时,  $n_r = 16r h_A^2 h_B^2$  如果不使用部分省略法,  $l_{11}$  的有效因复数为  $25r h_A^2 h_B^2$ 。这表示除  $A_1 \times B_1$  外,在其他交互作用可以省略的假定下,只用 5 列,就能保持 9 列全体的效率 64%。

对于四水平连续因素来说, 不仅希望  $A_1 \times B_1$  保持 64% 的效率, 且希望它提供完整的情报时, 可按图 46.2-14 进行处理。

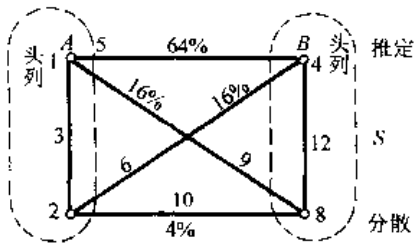


图 46.2-14  $A_1 \times B_1$  的情报分配比例

因而即使  $F$  值为 10 并以 1% 的风险率判定 5 列的波动是显著的, 那它只有 1/4 进入 6 列与 9 列。对于进入 6 列与 9 列的因素  $F$  值, 平均只有  $\pm 2.25$  的影响。即使把其他因素配列在 10 列中, 也只包含 4%, 所以是没有问题的, 但在 5 列的  $F$  值为 50 时, 就得修正 6 列与 9 列的效应。修正值为从 6 列与 9 列的效应估计值中减去 5 列效应的一半。

### 4.2 三水平系统的相互影响的多项式扩展

设  $A, B$  都是三水平的连续因素。例如  $A$  为温度,  $B$  为时间, 均等间隔的取三水平。  $A, B$  与实验值  $y$  的关系, 通常分解为正交多项式的成分, 即作如下的展开:

$$\begin{aligned}
 y = & m + a_1(A - \bar{A}) + a_2 \left[ (A - \bar{A})^2 - \frac{2}{3}h_A^2 \right] \\
 & + b_1(B - \bar{B}) + b_2 \left[ (B - \bar{B})^2 - \frac{2}{3}h_B^2 \right] \\
 & + l_{11}(A - \bar{A})(B - \bar{B}) \\
 & + l_{12}(A - \bar{A}) \left[ (B - \bar{B})^2 - \frac{2}{3}h_B^2 \right] \\
 & + l_{21}(B - \bar{B}) \left[ (A - \bar{A})^2 - \frac{2}{3}h_A^2 \right] \\
 & + l_{22} \left[ (A - \bar{A})^2 - \frac{2}{3}h_A^2 \right] \left[ (B - \bar{B})^2 - \frac{2}{3}h_B^2 \right]
 \end{aligned}$$

众所周知  $h_A, h_B$  为  $A, B$  的水平间隔, 只有  $A$  的一次项与二次项, 总称为  $A$  的主效应, 包括  $A, B$  两个因素的项  $(A - \bar{A})(B - \bar{B})$  称为  $A$  与  $B$  的交互作用。可是就主效应来说, 只考虑  $A, B$  得二次项以前的项, 而就交互作用来说却要考虑  $A^2B, A^2B^2$  这样的高次项, 这在一般情况下是不合适的。如果  $A, B$  是标牌、触媒之类的, 交互作用当然可以考虑四个项, 即 4 个自由度。但是  $A, B$  是连续变量, 如果能用多项式近似表现其效应, 就可以考虑交互作用的首项。

这就是说, 把  $y$  与  $A, B$  的关系表示为:

$$y = m + a_1(A - \bar{A}) + a_2 \left[ (A - \bar{A})^2 - \frac{2}{3}h_A^2 \right]$$

$$\begin{aligned}
 & + b_1(B - \bar{B}) + b_2 \left[ (B - \bar{B})^2 - \frac{2}{3}h_B^2 \right] \\
 & + l_{11}(A - \bar{A})(B - \bar{B})
 \end{aligned}$$

在上式中,  $A$  与  $B$  的交互作用只有一个。因而把  $A, B$  合计起来, 忽略三次以上的项时, 交互作用有一个自由度就够了。

### 4.3 三水平系统的相互影响部分忽略设计

连续因素间的交互作用, 往往可以省略三次以上的高次项。因而如果可以省略三次以上的项, 即使不全部确保 4 个的情报自由度, 也可以用  $A$  与  $B$  的积, 估计  $A$  与  $B$  的交互作用的首项。

例如我们用正交表  $L_{27}$  考虑图 46.2-15 所示的配列型式。唯有  $\odot$  的列表是互相混杂的。

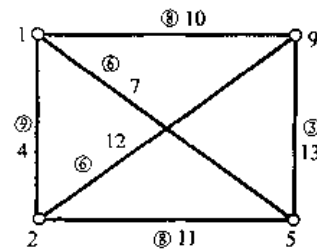


图 46.2-15 配列型式

本来, 对正交表  $L_{27}$  来说, 过去只取三个三水平因素的交互作用, 但在可以只考虑交互作用首项的情况下, 如图 46.2-15 所示, 可以考虑四个因素间的交互作用。例如设  $A, B, C, D$  为连续因素, 如果做如下对应:

- $A$  对应 1 列
- $B$  对应 2 列
- $C$  对应 5 列
- $D$  对应 9 列

则可以从 4 列求得  $A \times B$ , 从 7 列求得  $A \times C$ , 从 10 列求得  $A \times D$ , 从 11 列求得  $B \times C$ , 从 12 列求得  $B \times D$ , 从 13 列求得  $C \times D$ 。

例如对应于  $l_{11}(A - \bar{A})(B - \bar{B})$  项的估计值, 可以根据表 46.2-35 的  $A, B$  二元表, 按上述方法求出。其中正交表  $L_{27}$  有  $r$  次重复。

表 46.2-35  $A, B$  二元表与  $A$  列水平

$k = 3r$	$B_1$	$B_2$	$B_3$
$A_1$	1	2	3
$A_2$	3	1	2
$A_3$	2	3	1

有正交表的性质可知, 二元表中的 (1 之和) (2 之和) (3 之和) 是与  $A, B$  的主效应正交的。也就是对应于 4 列的 1、2、3 的和, 是  $A$  与  $B$  的交互作用的成分。但是我们只评价  $A \times B$  的首项, 所以对应于 (1 之和)、(2 之和)、(3 之和) 的  $l_{11}(A - \bar{A})(B - \bar{B})$  的

方程式, 可表示如下。

首先设  $A$ 、 $B$  等间隔取水平, 间隔分别为  $h_A$ 、 $h_B$ , 则有  $\bar{A} = A_2, B = B_2$ , 所以对  $A_1B_2, A_2B_1, A_2B_2, A_2B_3, A_3B_2$  来说,  $A_1 \times B_1$  项等于零。

$$9rm + [3r(-h_A)(-h_B) + 3r(h_A)(h_B)]l_{11} = (1\text{-之和})$$

$$(A_1B_1) \quad (A_3B_3)$$

$$9rm + 3r(h_A)(-h_B)l_{11} = (2\text{-之和})$$

$$(A_3B_2)$$

$$9rm + 3r(-h_A)(h_B)l_{11} = (3\text{-之和})$$

$$(A_1B_2)$$

把上式加以整理得:

$$9rm + 6rh_A h_B l_{11} = (1\text{-之和})$$

$$9rm - 3rh_A h_B l_{11} = (2\text{-之和})$$

$$9rm - 3rh_A h_B l_{11} = (3\text{-之和})$$

用最小二乘法解上式, 并令  $m$  与  $l_{11}$  正交, 则

$$\hat{m} = \frac{(1\text{-之和}) + (2\text{-之和}) + (3\text{-之和})}{27r}$$

$$\hat{l}_{11} = \frac{6(1\text{-之和}) + 3(2\text{-之和}) - 3(3\text{-之和})}{(6^2 + 3^2 - 3^2)rh_A h_B}$$

$$= \frac{2(1\text{-之和}) + (2\text{-之和}) + (3\text{-之和})}{18rh_A h_B}$$

$\hat{l}_{11}$  的显著性, 用下式检验

$$S_{A_1 \times B_1} = \frac{[2(1\text{-之和}) - (2\text{-之和}) - (3\text{-之和})]^2}{9r[2^2 + (-1)^2 + (-1)^2]}$$

$$= \frac{[2(1\text{-之和}) - (2\text{-之和}) - (3\text{-之和})]^2}{54r}$$

其自由度为 1。

$$\hat{l}_{11} \text{ 的方差为: } V_{\sigma}(\hat{l}_{11}) = \frac{\sigma^2}{6rh_A^2 h_B^2}$$

根据以上三式进行  $l_{11}$  估计、检验与确定估计值的置信界限。

$$\text{置信界限: } l_{11} = \hat{l}_{11} \pm \sqrt{\frac{F \times V_{\sigma}}{n_e}}$$

有效反复数  $n_e, n_e = 6rh_A^2 h_B^2$

尽管交互作用间存在混杂, 只要残留非混杂列, 当可以只考虑连续因素的 1 次  $\times$  1 次项目时, 就能对所有效应进行估计检验。混杂列是 3 列、6 列、8 列的波动合计。由此的方差分析表(见表 46.2-36)。

表 46.2-36 方差分析表

因素	$f$	$E(V)$
A	2	$\sigma_2^2 + \sigma_1^2 + 9\sigma_A^2$
B	2	$\sigma_2^2 + \sigma_1^2 + 9\sigma_B^2$
C	2	$\sigma_2^2 + \sigma_1^2 + 9\sigma_C^2$
D	2	$\sigma_2^2 + \sigma_1^2 + 9\sigma_D^2$
$A_i \times B_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_A^2 h_B^2 l_{11}^2$
$A_i \times C_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_A^2 h_C^2 l_{11}^2$

(续)

因素	$f$	$E(V)$
$A_i \times D_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_A^2 h_D^2 l_{11}^2$
$B_i \times C_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_B^2 h_C^2 l_{11}^2$
$B_i \times D_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_B^2 h_D^2 l_{11}^2$
$C_i \times D_i$	1	$\sigma_2^2 + \sigma_1^2 + 6rh_C^2 h_D^2 l_{11}^2$
混杂因素	6	$\sigma_2^2 + \sigma_1^2 + \text{交互作用}$
混杂因素	6	$\sigma_2^2 + \sigma_1^2$
混杂因素	$27(r-1)$	$\sigma_2^2$
计	$27r-1$	

#### 4.4 三水平系统和两水平系统之间的相互影响部分忽略设计

##### 4.4.1 配列

在三水平因素配列于二水平系统中, 可以采用拟因素法。

现在举例说明在二水平系统中如何求三水平间的交互作用。设有三水平因素  $A, B, C$ , 想求  $A \times B$ ; 现有二水平因素  $D, E, F, G$ , 由于想求  $D \times E$ , 所以用  $L_{16}$  作实验。但对于  $A \times B$  来说, 除  $A_1 \times B_1$  外, 其他项忽略。配列如表 46.2-37 所示。

表 46.2-37 用  $3^3 \times 2^4$  配列  $A \times B$  的例

No.	闲	A	B	A × B	E × D	C	D	F	G	E	e	e
		2,3	4,5	6	7	8,9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2	2
4	1	1	2	2	2	2	2	2	1	1	1	1
5	1	2	1	2	2	1	2	2	1	1	2	2
6	1	2	1	2	2	2	1	1	2	2	1	1
7	1	2	2	1	1	1	2	2	2	2	1	1
8	1	2	2	1	1	2	1	1	1	1	2	2
9	2	2	2	1	2	2	1	2	1	2	1	2
10	2	2	2	1	2	3	2	1	2	1	2	1
11	2	2	3	2	1	2	1	2	2	1	2	1
12	2	2	3	2	1	3	2	1	1	2	1	2
13	2	3	2	2	1	2	2	1	1	2	2	1
14	2	3	2	2	1	3	1	2	2	1	1	2
15	2	3	3	1	2	2	2	1	2	1	1	2
16	2	3	3	1	2	3	1	2	1	2	2	1

##### 4.4.2 分析法

设按  $L_{16}$  分别取了  $r$  个实验片, 就配列交互作用  $A \times B$  的 6 列来说, 如果能表出它与  $A, B$  二元素的关系, 如表 46.2-38 所示。

表 46.2-38

	$B_1$	$B_2$	$B_3$
$A_1$	1	2	—
$A_2$	2	11	2
$A_3$	—	2	1

表中的数字是 6 列的水平, A、B 是从值小的方向顺次开始标注水平号码。

在 6 列的 1 与 2 之差中, 由于除 A×B 外的其他因素全是正交的, 所以求得 1 与 2 之差后, 包含在其中的只有 A×B 的成分。

A×B 也只假定只取 A<sub>1</sub>×B<sub>1</sub>, 所以按上节所述的方法做同样的计算可得

$$\hat{l}_{11} = \frac{(1)-(2)}{4rh_A h_B}$$

$$S_{A \times B} = \frac{[(1)-(2)]^2}{16r}$$

$$V_{\mu}(\hat{l}_{11}) = \frac{\sigma^2}{rh_A^2 h_B^2}$$

在 7 列的情况下, A、B 二元表如表 46.2-39 所示。

表 46.2-39 A、B 二元表

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
A <sub>1</sub>	1	2	
A <sub>2</sub>	2	12	1
A <sub>3</sub>	—	1	2

所以可知 A×B 与 7 列正交。这就是说在 7 列的 1 与 2 之差中, 其他因素都未列入。尽管 A×B 与 D×C 都列入 7 列, 由于 A<sub>1</sub>×B<sub>1</sub> 因而能作 D×E 的评价。

S<sub>D</sub>, S<sub>E</sub>, S<sub>F</sub>, S<sub>D×E</sub>, S<sub>F</sub>, S<sub>G</sub>, 按通常方法计算。

$$S_A = S_{A_1} + S_{A_2}$$

$$= \frac{[1 \text{ 列是 } 1 \text{ 时, } 2 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

$$+ \frac{[1 \text{ 列是 } 2 \text{ 时, } 2 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

$$S_B = S_{B_1} + S_{B_2}$$

$$= \frac{[1 \text{ 列是 } 1 \text{ 时, } 4 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

$$+ \frac{[1 \text{ 列是 } 2 \text{ 时, } 4 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

$$S_C = S_{C_1} + S_{C_2}$$

$$= \frac{[1 \text{ 列是 } 1 \text{ 时, } 8 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

$$+ \frac{[1 \text{ 列是 } 2 \text{ 时, } 8 \text{ 列的 } (1 \text{ 之和}) \text{ 与 } (2 \text{ 之和}) \text{ 之差}]^2}{8r}$$

一次误差 S<sub>L<sub>1</sub></sub>, 可以用 14 列与 15 列的波动和来计算, 也可以由一次单位间的波动减去上述因素波动不求出。

二次误差 S<sub>L<sub>2</sub></sub>, 可以用总波动减去一次单位间的波动来求出, 也可以用 L<sub>16</sub> 的各号实验的重复波动 S 的和来求出。

由上述波动的计算结果, 做成方差分析表, 如表 46.2-40。

表 46.2 40 方差分析表

因 素	f	E(V)
空闲(1 列)	1	
A	A <sub>1</sub> , A <sub>2</sub>	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 4rσ <sub>A<sub>1</sub>, A<sub>2</sub></sub> <sup>2</sup>
	A <sub>2</sub> , A <sub>3</sub>	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 4rσ <sub>A<sub>2</sub>, A<sub>3</sub></sub> <sup>2</sup>
B	B <sub>1</sub> , B <sub>2</sub>	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> - 4rσ <sub>B<sub>1</sub>, B<sub>2</sub></sub> <sup>2</sup>
	B <sub>2</sub> , B <sub>3</sub>	σ <sub>2</sub> <sup>2</sup> - rσ <sub>1</sub> <sup>2</sup> - 4rσ <sub>B<sub>2</sub>, B<sub>3</sub></sub> <sup>2</sup>
A <sub>1</sub> ×B <sub>1</sub>	1	σ <sub>2</sub> <sup>2</sup> - rσ <sub>1</sub> <sup>2</sup> - rh <sub>A</sub> <sup>2</sup> h <sub>B</sub> <sup>2</sup> l <sub>11</sub> <sup>2</sup>
C	C <sub>1</sub> , C <sub>2</sub>	σ <sub>2</sub> <sup>2</sup> - rσ <sub>1</sub> <sup>2</sup> + 4rσ <sub>C<sub>1</sub>, C<sub>2</sub></sub> <sup>2</sup>
	C <sub>2</sub> , C <sub>3</sub>	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 4rσ <sub>C<sub>2</sub>, C<sub>3</sub></sub> <sup>2</sup>
D	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 8rσ <sub>D</sub> <sup>2</sup>
E	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 8rσ <sub>E</sub> <sup>2</sup>
D×E	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 4rσ <sub>D×E</sub> <sup>2</sup>
F	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 8rσ <sub>F</sub> <sup>2</sup>
G	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup> + 8rσ <sub>G</sub> <sup>2</sup>
e <sub>1</sub>	1	σ <sub>2</sub> <sup>2</sup> + rσ <sub>1</sub> <sup>2</sup>
e <sub>2</sub>	16(r-1)	σ <sub>2</sub> <sup>2</sup>
计	16r-1	

## 5 直和设计

### 5.1 直和设计的意义

当遇到因素较多而水平数又不相同的正交试验时, 如果用一张正交表来安排试验, 必然试验次数很多, 需要的周期很长, 同时也不可能在试验过程中及时发现问题和解决问题, 这时可把一个试验分阶段进行。先把一部分因素和水平安排在第一张正交表上进行试验, 若试验的结果达不到要求, 再利用这些结果提供的信息, 在第二张正交表上安排下一阶段的试验, 最后把两次试验结果进行综合分析, 得出结论, 称为直合法。

### 5.2 第一次实验方法

#### 5.2.1 第一次实验的配列

先利用减少因素、减少水平、同一列同时配列几个因素的办法, 作 3<sup>3</sup>×2<sup>3</sup> 型问题的第一次实验配列 (见表 46.2-41)。

就因素 G 来说, 由于把其水平固定在 G<sub>1</sub>, 第一次实验不能调查其效应。就因素 A 来说, 首先比较 A<sub>1</sub> 与 A<sub>2</sub>, 暂不考虑 A<sub>3</sub>。因素 B、C、D 分别进行两个水平的比较。因素 E 与 A 一样首先考虑比较 E<sub>1</sub> 与 E<sub>2</sub>, 暂不考虑 E<sub>3</sub>。因素 F 比较两个水平。对于温度 H, 由于 H<sub>2</sub> 与 H<sub>3</sub> 比 H<sub>1</sub> 重要, 所以首先比较 H<sub>2</sub> 与 H<sub>3</sub>, 如果 H<sub>3</sub> 比 H<sub>2</sub> 好, 就不比较 H<sub>1</sub> 与 H<sub>2</sub>, 而调查 H<sub>3</sub> 与 H<sub>1</sub>。

如此进行实验的结果, 问题就变为把七个二水平因素配列于 L<sub>8</sub>, 以求 A×B。

A<sub>1</sub> 与 A<sub>2</sub> 可配列于第一列, B<sub>1</sub> 与 B<sub>2</sub> 可配列于第二列, (A<sub>1</sub>, A<sub>2</sub>)×B 可配列于第三列。当剩下的五个

因素 C、D、(E<sub>1</sub>、E<sub>2</sub>)、F、(H<sub>1</sub>、H<sub>2</sub>) 不能原封不动地对应于 L<sub>8</sub> 剩下的四列, 这时就要用到复合法。根据情况虽然可以把五个因素中的一个暂不考虑处理, 但是用复合法还是有效的。

现在把 C 与 D 复合之后, 进行表头设计, 作如下对应:

- C<sub>1</sub>D<sub>1</sub>、C<sub>2</sub>D<sub>2</sub>→第四列
- F<sub>1</sub>、F<sub>2</sub>→第五列
- H<sub>1</sub>、H<sub>2</sub>→第五列
- 则 L<sub>8</sub> 的配列如下

表 46.2-41 第一次实验时, L<sub>8</sub> 中的因素与水平 (只提出标记下线的水平)

因素	水 平		
A	<u>A<sub>1</sub></u>	<u>A<sub>2</sub></u>	A <sub>3</sub>
B	<u>B<sub>1</sub></u>	<u>B<sub>2</sub></u>	
C	<u>C<sub>1</sub></u>	<u>C<sub>2</sub></u>	
D	<u>D<sub>1</sub></u>	<u>D<sub>2</sub></u>	E <sub>3</sub>
E	<u>E<sub>1</sub></u>	<u>E<sub>2</sub></u>	
F	<u>F<sub>1</sub></u>	<u>F<sub>2</sub></u>	
G	<u>G<sub>1</sub></u>	<u>G<sub>2</sub></u>	
H	<u>H<sub>1</sub></u>	<u>H<sub>2</sub></u>	<u>H<sub>3</sub></u>

表 46.2-42 第一次实验的 L<sub>8</sub> 配列

因素列 No.	A <sub>1</sub> A <sub>2</sub>	B	A <sub>1</sub> A <sub>2</sub> ×B	(C <sub>1</sub> D <sub>1</sub> ) (C <sub>2</sub> D <sub>2</sub> )	E <sub>1</sub> E <sub>2</sub>	F <sub>1</sub> F <sub>2</sub>	H <sub>1</sub> H <sub>2</sub>	数据小计			例 (减去虚拟平均值 70) 小 计		
	1	2	3	4	5	6	7	y <sub>11</sub>	y <sub>12</sub>	y <sub>i</sub>			
1	1	1	1	1	1	1	1	y <sub>11</sub>	y <sub>12</sub>	y <sub>i</sub>	-10	-7	-17
2	1	1	1	2	2	2	2	y <sub>21</sub>	y <sub>22</sub>	y <sub>i</sub>	-5	-5	-10
3	1	2	2	1	1	2	2	y <sub>31</sub>	y <sub>32</sub>	y <sub>i</sub>	-10	-11	-21
4	1	2	2	2	2	1	1	y <sub>41</sub>	y <sub>42</sub>	y <sub>i</sub>	-13	-10	-23
5	2	1	2	1	2	1	2	y <sub>51</sub>	y <sub>52</sub>	y <sub>i</sub>	2	3	5
6	2	1	2	2	1	2	1	y <sub>61</sub>	y <sub>62</sub>	y <sub>i</sub>	1	5	6
7	2	2	1	1	2	2	1	y <sub>71</sub>	y <sub>72</sub>	y <sub>i</sub>	-3	-6	-9
8	2	2	1	2	1	1	2	y <sub>81</sub>	y <sub>82</sub>	y <sub>i</sub>	2	3	5
G 固定为 G <sub>1</sub>								计	T				-64

5.2.2 数据分析: 方差分析

按照表 46.2-42 所示的配列进行 8 次实验 (八种不同条件)。就机械加工零件的实验来讲, 至少以不同条件各制造 3-4 个零件。就化学工业来讲, 以八种条件制造产品, 其每一种条件的产品构成一批, 一般每批要各做 2 个试片或者 2 个试样, 进行化学分析。如果各做两个试片, 就要考虑八种实验条件, 分别取测量值。对于中间装配产品和用设备试制生产的产品, 不必用两台设备分别制造, 应该使设备处于良好工作状态, 改变实验条件分别取数据。

这里有八种产品各做两个试片, 在同一条件下取

两个测量值, 这时数据的单纯重复数为 2。表 46.2-42 中第 i 号实验的两个重复数取为 y<sub>i1</sub>、y<sub>i2</sub>, 其合计为 y<sub>i</sub>。为了易于理解还给出数值列。为了简化计算, 数据是从原始数据减去虚拟平均值 70 后的代数值。方差分析如表 46.2-43 所示。

这时根据二次误差计算, 还有 A<sub>1</sub>A<sub>2</sub>×B、E (E<sub>1</sub>、E<sub>2</sub>)、E 三个因素的效应不显著, 所以可以认为一次误差很小。因为在这些不显著因素的 V 值中, 还包括 2 倍的一次误差, 因而在 e<sub>1</sub> 被否定的情况下, 通常把 V 值小的一次因素代入 e<sub>2</sub> 中, 合并后的误差波动 S<sub>e</sub> = 0.25 + 6.35 + 1.00 + S<sub>e2</sub> = 30.50, 其中自由度为 11, 然

后进行 F 检验。这意味着用正交表 L<sub>8</sub> 做实验, 各取 2

表 46.2-43 第一次实验的方差分析表

因 素	f	S	V	E(V)	F	ρ(%)
A(A <sub>1</sub> , A <sub>2</sub> )	1	380.25	380.25	σ <sub>A</sub> <sup>2</sup> - 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>A<sub>1</sub>A<sub>2</sub></sub> <sup>2</sup>	137.3	71.2
B	1	64.00	64.00	σ <sub>B</sub> <sup>2</sup> - 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>B</sub> <sup>2</sup>	23.1	11.6
(A <sub>1</sub> , A <sub>2</sub> )×B	1	0.25	0.25	σ <sub>B</sub> <sup>2</sup> + 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>A<sub>1</sub>A<sub>2</sub>×B</sub> <sup>2</sup>	—	
CD	1	25.00	25.00	σ <sub>C</sub> <sup>2</sup> + 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>CD</sub> <sup>2</sup>	9.0	4.2
E(E <sub>1</sub> , E <sub>2</sub> )	1	6.25	6.25	σ <sub>E</sub> <sup>2</sup> + 2σ <sub>e1</sub> <sup>2</sup> - 8σ <sub>E<sub>1</sub>E<sub>2</sub></sub> <sup>2</sup>	2.3	0.7
F	1	1.00	1.00	σ <sub>F</sub> <sup>2</sup> + 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>F</sub> <sup>2</sup>	—	
H(H <sub>1</sub> , H <sub>2</sub> )	1	30.25	30.25	σ <sub>H</sub> <sup>2</sup> - 2σ <sub>e1</sub> <sup>2</sup> + 8σ <sub>H<sub>1</sub>H<sub>2</sub></sub> <sup>2</sup>	10.9	5.2
e <sub>1</sub>	0	0.00		σ <sub>e1</sub> <sup>2</sup> + 2σ <sub>e1</sub> <sup>2</sup>	—	7.1
e <sub>2</sub>	8	23.00	2.87	σ <sub>e2</sub> <sup>2</sup>	(e)	
合并(e)	(11)	(30.50)	(2.77)			
计	15	530.00				100



个试片,共做 16 次实验,在精度上与各取 1 个试片的情况相同。

F 检验并非很重要。由于各因素的波动  $S$  大小,表示其影响程度大小,所以应该着眼于  $S$  的大小。如果像上述实验一样求得的误差方差  $V_e$  为 2.77,影响率  $p$  较波动  $S$  更实用一些。影响率的求算公式:

$$\rho_{A_1 A_2} = \frac{S_{A_1 A_2} - \text{该因素自由度} \times V_e}{\text{总波动 } S_T} \times 100$$

$$= \frac{380.25 - 1 \times 2.77}{530.0} \times 100 = 71.2\%$$

其他因素也一样。做八次实验,各取 2 个数据,共得 16 个数据,其中之所以不同,是由于故意改变  $A, B \dots$  的条件,存在测量误差以及其他原因,当它们在八次实验间发生变化时,总波动值就不等于零,对实验值的影响大,总波动值越大。在影响实验的所有原因中,把  $A$  改变为  $A_1$  与  $A_2$  的影响,占 71.2%,影响率就表示这一情况。

### 5.2.3 数据分析: 因素效应的估计

在数据分析中,比波动分解更重要的是因素效应的估计。因为  $A_1 A_2 \times B$  与  $F$  的主效应的方差比  $F$  都小于 2,所以略去。

置信界限的估计值用下式计算:

$$\pm \sqrt{\frac{F \times V_e}{n_e}} = \pm \sqrt{\frac{4.84 \times 2.77}{8}} = \pm 1.29$$

式中  $F$  是对应于分子自由度为 1,分母自由度为 11 (误差方差  $V_e$  的自由度为 11) 的  $F$  表的 5% 值;  $n_e$  是有效反复数,其估计值根据  $e_2$  反复几次的平均进行计算。本例各主效应都反复 8 次,如此求得估计值,如表 46.2-44 所示。如果把表 46.2-44 制成图表将会更清楚更直观。

表 46.2-44 估计

A	$A_1 = 70 + \frac{-71}{8} = 61.12 \pm 1.29$
	$A_2 = 70 - \frac{8}{7} = 70.88 \pm 1.29$
B	$B_1 = 68.00 \pm 1.29$
	$B_2 = 64.00 \pm 1.29$
CD	$C_1 D_1 = 70 + \frac{-42}{8} = 64.75 \pm 1.29$
	$C_2 D_2 = 70 + \frac{-22}{8} = 67.25 \pm 1.29$
E	$E_1 = 66.62 \pm 1.29$
	$E_2 = 65.38 \pm 1.29$
H	$H_2 = 64.62 \pm 1.29$
	$E_3 = 67.38 \pm 1.29$

## 5.3 直和实验方法

### 5.3.1 第二次实验的配列

根据第一次实验配列结果,设取大值为好,做下述判定。

A  $A_1$  与  $A_2$  的效应占因素效应的 71%,而且  $A_2$  比  $A_1$  大 9.76。因而想取更大的水平值  $A_3$  与  $A_2$  下一次比较。

B  $B$  的效应好,  $B_2$  的效应不好。但由于同  $A$  有交互作用,所以希望再比较一次  $B_1$  与  $B_2$  已确认的效应及其与  $A$  的交互作用。

C、D 把  $C, D$  复合配列的第四列有 4% 的影响率。在技术上认为是  $D$  的效应,但不能断言,所以希望做第二次实验配列,以分辨是  $C$  的效率还是  $D$  的效率,或者两者的效率。因而因素  $C, D$  都是取第一次实验的水平做第二次实验。

E  $E$  的效应不显著,影响率不到 1%,但作为数值,与预想的不同,  $E_1$  似乎效应显著。但由于水平可以连续的变化,所以希望因素  $E$  提出更小的水平值  $E_0$ ,比较  $E_0$  与  $E_1$ 。

F  $F_1, F_2$  都无显著效应。这在技术上可以理解的,所以在第二次实验中,可以把  $F$  的水平固定在成本低的  $F_1$  上。因此在第二次实验中  $F$  不需要作为配列因素提出。

G 在第一次实验时,  $G$  被固定在水平  $G_1$  上,在第二次实验时必须提出此因素,比较  $G$  与  $G_2$ 。

H  $H$  是连续因素,由于认为水平值大的  $H_3$  好,所以曾提出  $H_2$  与  $H_3$  比较,结果如预想的那样,  $H_3$  的效应显著。因而在第二次实验中,希望提出水平值大的  $H_4$ ,比较  $H_3$  与  $H_4$ 。

根据以上结果,第二次实验的因素与水平如表 46.2-45 所示。

表 46.2-45 第二次实验的因素与水平

因素	水平			
	$A_1$	$A_2$	$A_3$	$A_4$
A	$A_1$	$A_2$	$A_3$	
B	$B_1$	$B_2$		
C	$C_1$	$C_2$		
D	$D_1$	$D_2$		
E	$E_1$	$E_2$	$E_3$	
F	$F_1$	$F_2$		
G	$G_1$	$G_2$		
H	$H_1$	$H_2$	$H_3$	$H_4$

因而第二次实验希望求出七个二水平因素的效应,和  $A_2, A_3$  与  $B$  的交互作用。 $CD$  在第一次实验中是作为复合因素来求其效应的,这次为了分离开,不要忘记把水平组合颠倒过来,像表 46.2-46 那样配列。

表 46.2-46 第二次实验与数据

因素列 No.	A <sub>2</sub> A <sub>3</sub> 1	B 2	A <sub>2</sub> A <sub>1</sub> ×B 3	(C <sub>1</sub> D <sub>2</sub> ) (C <sub>2</sub> D <sub>1</sub> ) 4	E <sub>0</sub> E <sub>1</sub> 5	G <sub>1</sub> G <sub>2</sub> 6	H <sub>5</sub> H <sub>4</sub> 7	数据			例 (2 减去虚拟平均值 70)		
	1	2	3	4	5	6	7	小计	小计	小计	小计	小计	
1	1	1	1	1	1	1	1	y <sub>11</sub>	y <sub>12</sub>	y <sub>1</sub>	6	7	13
2	1	1	1	2	2	2	2	y <sub>21</sub>	y <sub>22</sub>	y <sub>2</sub>	6	5	11
3	1	2	2	1	1	2	2	y <sub>31</sub>	y <sub>32</sub>	y <sub>3</sub>	0	1	1
4	1	2	2	2	2	1	1	y <sub>41</sub>	y <sub>42</sub>	y <sub>4</sub>	-1	2	1
5	2	1	2	1	2	1	2	y <sub>51</sub>	y <sub>52</sub>	y <sub>5</sub>	11	11	22
6	2	1	2	2	1	2	1	y <sub>61</sub>	y <sub>62</sub>	y <sub>6</sub>	5	6	11
7	2	2	1	1	2	2	1	y <sub>71</sub>	y <sub>72</sub>	y <sub>7</sub>	6	4	10
8	2	2	1	2	1	1	2	y <sub>81</sub>	y <sub>82</sub>	y <sub>8</sub>	2	1	3
F 固定为 F <sub>1</sub>								(计)	72	(计)	72		

CD 是同第一次实验一样配列的，但并非必要的。

其结果如表 46.2-46 所示。第 i 号实验的数据取为 y<sub>i1</sub> 与 y<sub>i2</sub>，其和用 y<sub>i</sub> 表示。第一次实验记为 R<sub>1</sub>，第二次实验记为 R<sub>2</sub>。综合的方差分析如表 46.2-47 所示。把上述计算结果汇集起来，如表 46.2-48 所示。

5.3.2 综合分析：方差分析

表 46.2-3 的情况一样，设已得重复的 2 个数据。

表 46.2-47 综合的方差分析

计算式	计算例
$CF = \frac{(\text{全部合计})^2}{32} - \frac{(R_1 + R_2)^2}{32}$	$CF = \frac{(-64 + 72)^2}{32} - \frac{8^2}{32} = 2.0 (f=1)$
<p>A 的效应可分为 R<sub>1</sub> 的 A<sub>1</sub>, A<sub>2</sub> 的差 R<sub>2</sub> 的 A<sub>2</sub>, A<sub>1</sub> 的差来求</p> $S_{A_2A_3} = \frac{1}{16} (R_2A_2 - R_2A_3)^2$ $= \frac{1}{16} (y_1 + y_7 + y_3 + y_4 - y_2 - y_5 - y_6 - y_8)^2$	$S_{A_1A_2} = \frac{1}{16} (R_1A_1 - R_1A_2)^2$ $= 380.25 (f=1, \text{根据表 46.2-43})$ $S_{A_2A_3} = \frac{1}{16} (13 + 11 + 1 + 1 - 22 - 11 - 10 - 3)^2$ $= \frac{1}{16} (26 - 46)^2$ $= 25.00 (f=1)$
<p>B 的效应，进入了第一次实验，也进入了第二次实验中，所以要使用第一次实验与第二次实验的全部数据进行计算</p> $S_B = \frac{(B_1 - B_2)^2}{32}$ $= \frac{1}{32} (y_1 + y_2 + y_3 + y_6 - y_4 - y_5 - y_7 - y_8 - y_1 - y_2 - y_3 - y_6 - y_7 - y_8)^2$	$S_B = \frac{1}{32} [(-17 - 10 + 5 + 6 + 13 + 11 + 22 + 11) - (-21 - 23 - 9 - 5 + 1 + 1 - 10 + 3)]^2$ $= \frac{(41 + 33)^2}{32}$ $= 171.12 (f=1)$
<p>A × B 作如下分解之后来求出。</p> $S_{A \times B} = S_{A_1A_2 \times B} + S_{A_2A_3 \times B}$ <p>S<sub>A<sub>1</sub>A<sub>2</sub> × B</sub> = R<sub>1</sub> 的第三列波动</p> <p>S<sub>A<sub>2</sub>A<sub>3</sub> × B</sub> = R<sub>2</sub> 的第三列波动</p> $= \frac{1}{16} (y_1 + y_2 + y_7 + y_8 - y_3 - y_4 - y_5 - y_6)^2$	$S_{A_1A_2 \times B} = 0.25 (f=1, \text{根据表 46.2-43})$ $S_{A_2A_3 \times B} = \frac{1}{16} (R_2 \text{ 的第三列 1 之和} - R_2 \text{ 的第三列 2 之和})^2$ $= \frac{1}{16} (37 - 35)^2$ $= 0.25 (f=1)$
<p>S<sub>C</sub> 的计算，可以不考虑 C 与 D 的复合，使用 R<sub>1</sub> 与 R<sub>2</sub> 的数据。</p> $S_C = \frac{1}{32} (C_1 - C_2)^2$ $= \frac{1}{32} [(y_1 + y_3 + y_5 + y_7 + y_1 - y_3 + y_5 + y_7) - (y_2 + y_4 + y_6 + y_8 + y_2 + y_4 + y_6 + y_8)]^2$	$S_C = \frac{1}{32} [(-17 - 21 + 5 - 9 + 13 + 1 + 22 - 10) - (-10 - 23 + 6 + 5 + 11 + 1 + 11 + 3)]^2$ $= 0 (f=1)$
<p>S<sub>D</sub> 的计算，与 S<sub>C</sub> 一样，不考虑 C<sub>1</sub> 而根据 R<sub>1</sub> 与 R<sub>2</sub> 两次实验的数据</p> $S_D = \frac{(D_1 - D_2)^2}{32}$	$S_D = \frac{1}{32} [(-17 - 21 + 5 - 9 + 11 + 1 + 11 + 3) - (-10 - 23 + 6 + 5 + 13 - 1 + 22 + 10)]^2$ $= 50.0 (f=1)$

(续)

计算式	计算例
$S_E = S_{E_1 E_2} + S_{E_0 E_1}$ $S_{E_1 E_2} = \frac{1}{16} (R_1 E_1 - R_1 E_2)^2$ $S_{E_0 E_1} = \frac{1}{16} (R_2 E_0 - R_2 E_1)^2$ $= \frac{1}{16} [(y_1 + y_3 + y_6 - y_8) - (y_2 + y_4 - y_5 + y_7)]^2$	$S_{E_1 E_2} = 6.25, (f=1, \text{根据表 46.2-44})$ $S_{E_0 E_1} = \frac{1}{16} [(13+1+11+3) - (11+1+22+10)]^2$ $= 16.00 (f=1)$
$S_F$ 只根据 $R_1$ 求算 $S_F = \frac{1}{16} (F_1 - F_2)^2$	$S_F = 1.00 (f=1, \text{根据表 46.2-44})$
$S_G$ 只根据 $R_2$ 求算 $S_G = \frac{1}{16} [y_1 + y_4 + y_5 + y_8] - (y_2 + y_3 + y_6 + y_7)]^2$	$S_G = \frac{1}{16} [(13+1+22+3) - (11+1+11+10)]^2$ $= 2.25 (f=1)$
$S_H = S_{H_2 H_3} + S_{H_3 H_4}$ $S_{H_2 H_3} = \frac{1}{16} (R_1 H_2 - R_1 H_3)^2$ $S_{H_3 H_4} = \frac{1}{16} (R_2 H_3 - R_2 H_4)^2$ $= \frac{1}{16} [(y_1 + y_4 + y_6 - y_7) - (y_2 + y_3 + y_5 + y_8)]^2$	$S_{H_2 H_3} = 0.25 (f=1, \text{根据表 46.2-44})$ $S_{H_3 H_4} = \frac{1}{16} [(13+1) + 11 + 10] - (11+1+22+3)]^2$ $= 0.25 (f=1)$
<p>对应于 <math>R_1</math> 与 <math>R_2</math> 之间的差异的波动,是需要计算的。在 <math>S_R</math> 中,有许多因素混杂着</p> $S_R = \frac{1}{32} (R_1 - R_2)^2$ $= \frac{1}{32} [(y_1 + \dots + y_8) - (y_1 + \dots + y_8)]^2$	$S_R = \frac{1}{32} (-64 - 72)^2$ $= 578 (f=1)$
<p>一次误差波动 <math>S_{e_1}</math> 等于</p> $S_{e_1} = \frac{1}{32} (R_1 B_1 + R_2 B_2 - R_1 B_2 - R_2 B_1)^2$	$S_{e_1} = \frac{1}{32} (-1 - 9)^2$ $= 3.12 (f=1)$
$S_{e_2} = \frac{1}{2} [(y_{11} - y_{12})^2 - (y_{21} - y_{22})^2 + \dots + (y_{81} + y_{82})^2$ $+ (y_{11} - y_{12})^2 + \dots + (y_{81} - y_{82})^2]$ (此结果用于验算)	$S_{e_2} = \frac{1}{2} [(-10+7)^2 + (-5+5)^2 + \dots + (2-3)^2 + (6-7)^2$ $+ \dots + (2-1)^2]$ $= 32.00 (f=16)$

表 46.2-48 综合的方差分析表

因 数	f	S	V	F	$\rho(\%)$	E(V)
R	1	578.00	578.00	340.00	44.5	$\sigma_2^2 + 2\sigma_1^2 + \text{交错因素}$
A	$\begin{cases} A_1 & A_2 \\ A_2 & A_3 \end{cases}$	380.25	380.25	223.70	29.2	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{A_1 A_2}^2$
	$\begin{cases} A_2 & A_3 \end{cases}$	25.00	25.00	14.70	1.8	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{A_2 A_3}^2$
B	1	171.12	171.12	100.70	13.1	$\sigma_2^2 + 2\sigma_1^2 + 16\sigma_B^2$
A×B	$\begin{cases} A_1 A_2 \times B \\ A_2 A_3 \times B \end{cases}$	0.25	0.25	—	—	$\sigma_2^2 + 2\sigma_1^2 + 4\sigma_{A_1 A_2 \times B}^2$
	$\begin{cases} A_2 A_3 \times B \end{cases}$	0.25	0.25	—	—	$\sigma_2^2 + 2\sigma_1^2 + 4\sigma_{A_2 A_3 \times B}^2$
C	1	0.00	0.00	—	—	$\sigma_2^2 + 2\sigma_1^2 + 16\sigma_C^2$
D	1	50.00	50.00	29.40	3.7	$\sigma_2^2 + 2\sigma_1^2 + 16\sigma_D^2$
E	$\begin{cases} E_1 & E_2 \\ E_0 & E_1 \end{cases}$	6.25	6.25	3.7△	0.4	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{E_1 E_2}^2$
	$\begin{cases} E_0 & E_1 \end{cases}$	16.00	16.00	9.4	1.1	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{E_0 E_1}^2$
F	1	1.00	1.00	—	—	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_F^2$
G	1	2.25	2.25	1.3	—	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_G^2$
H	$\begin{cases} H_2 & H_3 \\ H_3 & H_4 \end{cases}$	30.25	30.25	17.80	2.2	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{H_2 H_3}^2$
	$\begin{cases} H_3 & H_4 \end{cases}$	0.25	0.25	—	—	$\sigma_2^2 + 2\sigma_1^2 + 8\sigma_{H_3 H_4}^2$
e <sub>1</sub>	1	3.12	3.12	(V <sub>e<sub>1</sub></sub> /V <sub>e<sub>2</sub></sub> = 1.56)		$\sigma_2^2 + 2\sigma_1^2$
e <sub>2</sub>	16	32.00	2.00			$\sigma_2^2$
合并(e)	(23)	(39.12)	(1.70)		4.0	
计	31	1296.00			100.00	

首先,用  $e_2$  检验  $e_1$  得知  $e_1$  无显著效应,同时有几个因素用  $e_2$  检验也无显著效应。因而把这些无显著效应的因素  $A \times B, C, F, G, (H, H_1), e_1$  并入  $e_2$  中,可求新的波动误差  $S_e$ 。

$$S_e = 0.25 + 0.25 + 0.00 + 1.00 + 2.25 \\ + 3.12 + 32.00 = 38.87$$

$$f = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 16 = 23$$

$A_1$  与  $A_2$  之差大,  $A_2$  与  $A_3$  之差不大,这可从  $A$  的效应分解后的影响率看出来。同时  $C$  与  $D$  的复合效应实际上是  $D$  的效应,这也可从  $C$  无显著效应,  $D$  的影响率有 3.7% 看出来。其他因素的效应只占全部的 4%,所以可知已提出的因素对效应的影响很大

可以看到  $R$  具有最大的影响率。但在  $R$  的效应中,  $A, B, E, F, G, H$  的效应同第一与第二次实验条件之差,是全部互相混杂的。如果详细计算,从  $R$  的效应中修正  $A, A \times B, E, F, G$  的效应,就能计算第一次实验与第二次实验之差的纯波动,但最好不用这种非正交性情况。

### 5.3.3 综合分析:因素效应的估计

就因素  $A$  来说,当  $A_1$  与  $A_2$  之间,  $A_2$  与  $A_3$  之间,都存在差异时,  $A$  的效应图的绘制,应描述以下三点。 $A_1$  与  $A_3$  的比较以  $A_2$  为媒介。

设第一次实验的  $A_1, A_2$  的平均为  $\overline{A_1 R_1}, \overline{A_2 R_1}$ ,则这个值就成为  $A_1$  与  $A_2$  的公平比较。

但同  $A_3$  的比较,有欠公平,所以,根据第二次实验的  $A_2, A_3$  的平均为  $\overline{A_2 R_2}, \overline{A_3 R_2}$ ,但由于  $A$  的条件相同,  $\overline{A_2 R_1}$  与  $\overline{A_2 R_2}$  两者就成为第一次实验和第二次实验的估计值。因此通过第一次实验与第二次实验的全体,估计  $A$  的效应,或绘制效应图时,必须计算修正量  $W$ ,以修正数据。为此可用下述公式:

$$W = \text{修正值} = \frac{1}{2} (\overline{A_2 R_2} - \overline{A_2 R_1}) \\ = \frac{1}{16} [(y_1 + y_2 + y_3 + y_4) - (y_5 + y_6 + y_7 + y_8)]$$

$$A_1 \text{ 的平均} = \overline{A_1 R_1} + W \\ = \frac{1}{8} (y_1 + y_2 + y_3 + y_4) + W$$

$$A_2 \text{ 的平均} = \overline{A_2 R_2} + \overline{A_2 R_1} \\ = \frac{1}{16} [(y_1 + y_2 + y_3 + y_4) \\ + (y_5 + y_6 + y_7 + y_8)]$$

$$A_3 \text{ 的平均} = \overline{A_3 R_2} - W \\ = \frac{1}{8} (y_5 + y_6 + y_7 + y_8) - W$$

在计算置信界限时,由于把重点放在  $A_1, A_2, A_3$  间的相对差上,所以任意反复数为 8,即可以用下列公

式计算置信界限:

$$\pm \sqrt{\frac{F_{23} \times V_e}{8}}$$

当然对于  $B, C, D$  的效应估计值来讲,有效反复数为 16,以下说明具体的计算内容。

$A$  的效应估计

$$W = \frac{1}{2} (\overline{A_2 R_2} - \overline{A_2 R_1}) = \frac{1}{2} \left( \frac{26}{8} - \frac{7}{8} \right) = 1.2$$

$$\overline{A_1} = 70 + \frac{-71}{8} + W = 70 + (-8.9) + 1.2 \\ = 62.3 \pm 0.9$$

$$\overline{A_2} = 70 - \frac{1}{16} (26 - 7) = 72.1 \pm 0.9$$

$$\overline{A_3} = 70 + \frac{46}{8} - W = 74.6 \pm 0.9$$

置信界限是按下述方法求出的,就  $A_2$  的平均来说,它的形式上是 16 个数的平均,实质上只是  $\overline{A_1 R_1}$  与  $\overline{A_2 R_1}$  的比较精度与  $\overline{A_2 R_2}$  与  $\overline{A_1 R_2}$  的比较精度,所以与  $A_1, A_3$  一样,具有有效反复数  $n_e$  也等于 8。因此置信界限等于

$$\pm \sqrt{\frac{F_{23} \times V_e}{8}} = \pm \sqrt{\frac{4.28 \times 1.70}{8}} = \pm 0.91 \approx \pm 0.9$$

对于  $B, C, D$ ,有效反复数为 16,置信界限为上式的  $1/\sqrt{2}$ ,即为  $\pm \frac{0.91}{\sqrt{2}} = \pm 0.60$

$B$  的效应估计

$$\overline{B_1} = \frac{1}{16} (B_1) = 70 + \frac{41}{16} = 72.6 \pm 0.6$$

$$\overline{B_2} = \frac{1}{16} (B_2) = 70 + \frac{-33}{16} = 67.9 \pm 0.6$$

$D$  的效应估计

$$\overline{D_1} = 70 + \frac{-16}{16} = 69.0 \pm 0.6$$

$$\overline{D_2} = 70 + \frac{24}{16} = 71.5 \pm 0.6$$

$E$  的效应估计

$$W = \frac{1}{2} (\overline{R_2 E_1} - \overline{R_1 E_1}) = \frac{1}{16} (44 + 27) \\ = 4.4$$

$$\overline{E_2} = \overline{R_2 E_2} \quad W = 70 + \frac{28}{8} - 4.4 \\ = 69.1 \pm 0.9$$

$$\overline{E_1} = (\overline{R_2 E_1} + \overline{R_1 E_1}) = 70 + \frac{1}{16} (44 - 27) \\ = 71.1 \pm 0.9$$

$$\overline{E_2} = \overline{R_1 E_2} + W = 70 + \frac{-37}{8} + 4.4 \\ = 69.8 \pm 0.9$$

$G$  的效应估计

G 的效应本来不显著且影响率低。本无需进行估计,这里进行估计为了提供参考数据,因为在  $R_1$  的情况下,G 被固定在  $G_1$  上,

$$\begin{aligned} \bar{G}_1 &= \bar{T} + \frac{1}{2}(\bar{R}_2 G_1 - \bar{R}_2) \\ &= 70 + \frac{8}{32} + \frac{1}{2} \left( \frac{39}{8} - \frac{72}{16} \right) \\ &= 70.4 \pm 0.9 \end{aligned}$$

$$\begin{aligned} \bar{G}_2 &= \bar{T} + \frac{3}{2}(\bar{R}_2 G_2 - \bar{R}_2) \\ &= 70 + \frac{8}{32} + \frac{3}{2} \left( \frac{33}{8} - \frac{72}{16} \right) \\ &= 69.7 \pm 0.9 \end{aligned}$$

H 的效应估计

$$\bar{H}_2 = 68.1 \pm 0.9$$

$$\bar{H}_3 = 70.9 \pm 0.9$$

$$\bar{H}_4 = 71.1 \pm 0.9$$

用效应图 46.2-16 表示这些结果。

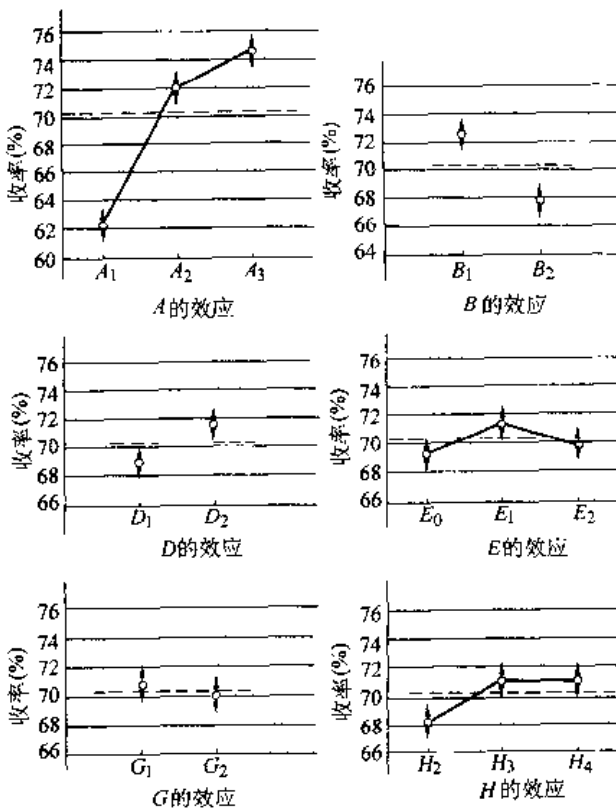


图 46.2-16 因素效应图

### 5.3.4 最宜条件的决定——工序的平均估计

最后对平均的最大组合,进行估计。这里全部实验值的平均为:

$$\bar{T} = 70.0 + \frac{8}{32} = 70.25$$

把  $\bar{T}$  代入图 46.2-16 的各图中,只要读取此平均值

与各因素水平的图示值之差,把 70.25 按代数和和计算就行了。此法也可以用来选择实验的水平以外的中间水平。

但是这里在实际上是考虑实验水平的最宜组合。由图 46.2-16 可知,因此 A 以  $A_3$  最好;因素 B 以  $B_1$  最好;因此 C 的任一水平都好,可以不作考虑;因素 D 以  $D_2$  最好;因素 E 以  $E_1$  最好;因素 F 可以不作考虑;因素 G 以  $G_1$  最好;因素 H 选  $H_4$ 。

设  $A_3 B_1 D_2 E_1 G_1 H_4$  的总体平均估计值为  $\hat{\mu}$ ,则在最宜条件下确认实验时,实现  $x$  的预测值等于

$$\begin{aligned} x &= \bar{A}_3 + \bar{B}_1 + \bar{D}_2 + \bar{E}_1 + \bar{G}_1 + \bar{H}_4 - 5\bar{T} \\ &\pm \sqrt{4.28 \times 1.70 \times \left( \frac{1}{n_e} + 1 \right)} \end{aligned}$$

代入实际数据得:

$$\begin{aligned} x &= 74.6 + 72.6 + 71.5 + 71.1 + 70.4 + 71.1 \\ &\quad - 5 \times 70.25 = 80.1 \pm 3.3 \end{aligned}$$

有效反复数  $n_e$  的计算

$$\begin{aligned} n_e &= \frac{\text{总实验数}}{\text{对 } \hat{\mu} \text{ 不能忽视的因素的实际有效自由度之和}} \\ &= \frac{32}{1(m) + 5(A) + 1(B) + 1(D) + 1(E) + 0.5(G) + 5(H)} \\ &= \frac{32}{14.5} \end{aligned}$$

式中单纯由一方进行实验的  $A_3$  与  $H_3$ ,其实际有效自由度按 5 计算,由两方进行实验的  $E_1$ ,其实际有效自由度取 1;对仅用一方进行实验的 G 来说,其实际有效自由度在  $G_1$  时 0.5,在  $G_2$  时取 4.5。

用直和法进行实验的注意事项:

- (1) 决定实验目的
- (2) 可行手段

当实验目的决定后,为实现目的要考虑成本上设备上所有可行的手段。不管是否对实现目的有用,都要事先把所有手段全部提出来,即使有用性小也应当提出。

因素选择上不良原因的位置和消除不良原因的手段可能是两回事。例如在某种化学反应中,其混合液的过滤工序有 5% 的损耗,这时虽然已知发生损失的原因在于过滤工序,然而如果只把改进过滤工序作为消除不良原因的手段,可以说是技术上有些一无所知。原因在于通过过滤的粘度与温度的改变,或加入适量的分散剂,用现有的过滤机也许能使损失消除。消除不良或者损失的方法是多种多样的,只要不是减少损失对最有效最便宜的方法,就不能说是最佳技术条件。

再者仅仅改变某一特性,即使能取得效果,通常也必须考虑全部特性与成本。例如用原油精炼高级润滑剂,总希望提高收率,这时提高收率,只要增加溶剂就

行了。但这样做如果使得另一特性——凝固温度恶化而产生不良品,就是一种不正确的方法。如果增加溶剂可以提高收率,那么如果能找到不影响收率却能改进凝固温度的因素就可行了,也就是说,既对收率没有太大的作用,又能够改进凝固温度的因素。除此之外,在做实验时,还应考虑对非目的特性有影响的因素,因而在实验中应该提出的因素往往很多。

(3) 实验配列

因素与水平选齐后,要编制因素水平表,再从其中选出第一次实验用的因素。这时应该把准备时间不需太长的因素和水平多的因素提出,列入第一次实验的正交表中,应该大胆运用复合法,即同一列安排 2 个以上的因素的方法。采用复合法应尽可能地掌握因素的定性趋势。

(4) 数据测量

例如在制造八种产品时,重要的问题是与该产品的实验条件有关的问题。数据的测量的方法有:八种产品各制造八个或九个试样,分别按正交表的各种实验条件各测一个数据;八次实验的每一次都制造几个试样或者产品,进行同一条件下的重复测量;一般希望在改变使用上进行测量一个数据。

(5) 数据分析

例如对于前节  $L_9$  实验经过两次直和实验的方差分析计算。除此重要的是因素效应图分析,根据因素效应图,考虑水平差所引起的成本差异与因素效应两个方面,找出最宜条件。对应于最宜条件的各种特性值估计式很多。有时可以把数据进行分组,运用计数分类值的分析方法,即累计法进行计算。

6 部分扩展设计

用某种正交表配列时,由于因素过少、同时又不想用大的正交表来进行实验时,想用实验所得数据的一部分进行后续工序研究时,设备装置有某种限制条件时,均可以使用部分扩展法。

6.1  $L_9$  的部分扩展

设因素均 A、B、C、D、E、F 为二水平。把因素 A、B、C 配列于  $L_9$  正交表的第二、三、四列之后,如下表所示使 No. 1、No. 2、No. 3 实验扩大为 3 倍。可以这么认为,把原正交表  $L_9$  的第一列、第一水平的 No. 1、No. 2、No. 3 实验扩大为 3 倍,而变为实施部分扩大的正交表的第一列。因而加上第二、三、四列就形成新的五、六、七列。对应于部分扩大的列——第五、六、七列的因素 D、E、F,按 No. 10~No. 15 实验固定其水平, D、E、F 的效应要使用 No. 1~No. 9 实验数据求算, A、B、C 的效应要使用 No. 10~No. 15 是把应该作三次的

相同实验,省略为一次,故取数据 3 倍值。

波动值的计算,如下:

$$CF = \frac{288^2}{27} = 3072$$

$$S_T = 5^2 + 7^2 + \dots + 8^2 - 3(9^2 + 13^2 + \dots + 10^2) - CF = 328 \quad (f=14)$$

$$S_{PH} = \frac{96^2 + 90^2 + 102^2}{9} - CF = 8 \quad (f=2)$$

$$S_A = \frac{(21+27+45)^2 + \dots + (42+24+30)^2}{9} - CF = 2 \quad (f=2)$$

$$S_B = 168$$

$$S_C = 8$$

$$S_D = \frac{(5+8+16)^2 + \dots + (9+9+8)^2}{3}$$

$$= \frac{(21+33+42)^2}{9}$$

$$= 29 \quad (f=2)$$

$$S_E = 1 \quad (f=2)$$

$$S_F = 113 \quad (f=2)$$

$$S_e = 0 \quad (f=0)$$

由上述结果,得到表 46.2-49 的方差分析表。

表 46.2-49 方差分析表

因素	f	S	V	E(V)	修正系数	V'
赋闲列	2	8	4.0	$\frac{7}{3}\sigma^2 + 9\sigma_{PH}^2$	3/7	1.7
A	2	2	1.0	$\frac{7}{3}\sigma^2 + 9\sigma_A^2$	3/7	0.4
B	2	168	84.0	$\frac{7}{3}\sigma^2 + 9\sigma_B^2$	3/7	36.0
C	2	8	4.0	$\frac{7}{3}\sigma^2 + 9\sigma_C^2$	3/7	1.7
D	2	29	14.5	$\frac{7}{3}\sigma^2 + 9\sigma_D^2$	1	14.5
E	2	1	0.5	$\frac{7}{3}\sigma^2 + 9\sigma_E^2$	1	0.5
F	2	133	56.5	$\frac{7}{3}\sigma^2 + 9\sigma_F^2$	1	56.5
e	0	0		(V <sup>2</sup> )		
0号并入(e)	(8)	(4.3)				(0.54)
计	14		327			

方差期望值按下述规则计算。

部分扩大法的方差期望值求法 过剩的误差方差为  $(27-15)\sigma^2 = 12\sigma^2$ ,被扩大前的正交表  $L_9$  的全部自由度平均分配。因而,闲、A、B、C 的波动为

$$2\sigma^2 + \frac{12}{9}\sigma^2 \times 2 = \frac{14}{3}\sigma^2$$

$$m(\text{一般平均}) \text{ 为 } \sigma^2 + \frac{4}{3}\sigma^2 = \frac{7}{3}\sigma^2$$

因素效应部分,与做 27 次试验的情况相同。修正后的方差  $V'$ ,由  $V$  乘修正系数求出。把闲、A、C、E 合并之后,便得到误差(e)。

显著性因素按下述方法估计。

$$\bar{T} = \frac{288}{27} = 10.7 \pm 0.5$$

$$\bar{B}_1 = \frac{72}{9} = 8.0 \pm 0.9$$

$$\bar{B}_2 = \frac{90}{9} = 10.0 \pm 0.9$$

$$\bar{B}_3 = \frac{126}{9} = 14.0 \pm 0.9$$

$$\bar{D}_1 = \frac{39}{3} = 13.0 \pm 1.0$$

$$\bar{D}_2 = \frac{31}{3} = 10.3 \pm 1.0$$

$$\bar{D}_3 = \frac{26}{3} = 8.7 \pm 1.0$$

$$\bar{F}_1 = \frac{19}{3} = 6.3 \pm 1.0$$

$$\bar{F}_2 = \frac{32}{3} = 10.7 \pm 1.0$$

$$\bar{F}_3 = \frac{45}{3} = 15.0 \pm 1.0$$

$\bar{T}$ 、 $B$  的有效反复数: 方差  $V$  的期望值中的有效反复数可以用修正系数乘 27 与 9 计算得。例如对于  $B$  而言

$$\pm \sqrt{\frac{5.32 \times 0.54}{9 \times 3/7}} = \pm 0.9$$

$D$ 、 $F$  的有效反复数为 3。

数值最大的组合  $B_3D_1F_3$  的工序平均, 按下述方法求算。就配列于部分扩大的  $D$ 、 $F$  来说, 第一水平做了 21 次实验, 第二、第三水平都只做了 3 次实验, 所以把重叠部分加上。因此平均工序  $\hat{\mu}$  等于

$$\begin{aligned} \hat{\mu} &= \bar{T} + (\bar{B}_3 - \bar{T}) + \frac{6\bar{D}_1 - 3\bar{D}_2 - 3\bar{D}_3}{27} \\ &+ \frac{24\bar{F}_3 - 21\bar{F}_1 - 3\bar{F}_2}{27} \pm \sqrt{F \times V_e \left( \frac{1}{n_e} + 1 \right)} \\ &= 14.0 + \frac{2 \times 13 - 10.3 - 8.7}{9} \\ &+ \frac{8 \times 15 - 7 \times 6.3 - 10.7}{9} \\ &\pm \sqrt{5.32 \times 0.54 \times \left( \frac{61}{81} + 1 \right)} \\ &= 22.0 \pm 2.2 \end{aligned}$$

一般来说,  $B_1$ 、 $B_2$ 、 $B_3$  的互相比, 能得到相等次数的比较情报, 但在实验数据的总平均  $\bar{T}$  中,  $r_1$ 、 $r_2$ 、 $r_3$  次的的数据被平均时, 如选水平  $B_1$ , 则加  $\bar{T}$  的效应, 可用下式给出。由于各项是独立的, 所以  $n_e$  是根据单位数定义求出的。

$$\text{选择 } B_1 \text{ 的效应} = \frac{(r_2 + r_3)\bar{B}_1 - r_2\bar{B}_2 - r_3\bar{B}_3}{r_1 + r_2 + r_3}$$

如果求有效反复数感到困难时, 也可以用下式求算得各个预测值的近似置信界限:

$$+ \sqrt{8 \times V_e} = \pm \sqrt{8 \times 0.54} = \pm 2.1$$

正交表的部分扩大, 并不重要, 因为可以使用正交表  $L_{16}$ 。但在用三水平十五个因素, 想求三个交互作用时, 如果第一列赋闲, 把其第一水平进行部分扩大, 则用 45 次实验便可求得全部情报。同时按正交表  $L_{27}$  制造毛坯, 把其中的误差列的一个水平进行部分扩大, 其后续的机械加工便能进行三水平九因素的实验。至于剩下的毛坯的 18 次实验, 把机械加工条件固定, 便可以得到共计 45 种加工产品。而且用毛坯实验配列的因素, 能够求机械加工后的效应。

### 6.2 $L_8$ 的部分扩展

设有某种化学反应的生产试制实验。根据某一次反应, 提出了二水平因素  $A$ 、 $B$ 。一次反应所得一批产品做三等分, 投入二次反应的三个罐  $R_1$ 、 $R_2$ 、 $R_3$  中进行二次反应。根据二次反应, 提出了三水平因素  $C$  和二水平的四个因素  $D$ 、 $E$ 、 $F$ 、 $G$ , 只取主效应进行了正交表  $L_8$  的部分扩大配列。这种情况下的部分扩大配列, 首先把罐分为两组:  $R'_1(R_1R_2)$  与  $R'_2(R_3)$ , 像二水平那样配列于正交表  $L_8$ 。 $A$ 、 $B$  配列于第二、三列。 $R'$  配列于第四列,  $D$  配列于第五列,  $C$  配列于第六、七列, 第一列可赋闲。可是第四列的第一水平配列着  $R'_1$ , 但其中由两个罐  $R_1$ 、 $R_2$ 。用  $R_1$ 、 $R_2$  作相同的实验时, 这就是  $R_3$  的部分追加, 至于  $E$ 、 $F$ 、 $G$ , 两个罐自然是用来做不同的实验。因而部分扩大也是部分追加法的一般化。为使第四列第一水平进行部分扩大, 就要把第四列的第一水平的四种实验考虑为四个水平, 配列于正交表  $L_8$  的第一、二、三列的四个水平上。如此这样就能产生四个新列——第四、五、六、七列(应为八、九、十、十一), 把  $R_1$ 、 $R_2$ 、 $E$ 、 $F$ 、 $G$  配列于其中。对应于第四列的第二水平,  $E$ 、 $F$ 、 $G$  都把水平固定在第一水平上。因而配列如表 46.2-50 所示:

表 46.2-50 配列

	一次反应			二次反应							数据	
	闲 1	A 2	B 3	R 4	R 8	D 5	C 67	E 9	F 10	G 11	(- 60)	小 计
1	1	1	1	1 1 2	1 2 3	1 1 2	1 1 2	1 2 1	1 2 1	1 2 1	0 9 4	9 8
2	1	2	2	1 1 2	1 2 3	1 1 2	2 2 1	1 2 1	2 1 1	2 1 1	18 8 4	26 8 8
3	2	1	2	1 1 2	1 2 3	2 2 1	2 2 3	2 2 1	1 2 1	2 1 1	18 17 6	25 12
4	2	2	1	1 1 2	1 2 3	2 2 1	3 2 3	2 1 1	2 1 1	1 2 1	13 1 3	14 6
计 108												

数据如数据栏所示,是收率减去虚拟平均后的结果,因此进而求解下列问题。

- 1) 方差分析;
- 2) 显著性因素估计;
- 3) 最适宜工序的决定与工序平均估计。

### 6.2.1 方差分析

$R_1$ 、 $R_2$  的数据相加的,  $R_3$  的数据按 2 倍计算并求出小计。赋闲列因素  $C$  的波动计算,像做 16 次实验一样进行;  $R_1$ 、 $R_2$ 、 $E$ 、 $F$ 、 $G$  的波动计算,要扣除  $R_3$  的数据。

$$CF = \frac{108^2}{16} = 729.0$$

$$S_T = 0^2 + 9^2 + 2 \times 4^2 + A + 13^2 + 1^2 + 2 \times 3^2 - CF = 417.0$$

$$S_{m'} = \frac{1}{16} (51 - 57)^2 = 2.25 = 2.3$$

$$S_A = \frac{1}{16} (54 - 54)^2 = 0.0$$

$$S_B = 72.7$$

$$S_{R'} = \frac{1}{16} (74 - 34)^2 = 100$$

$$S_D = \frac{1}{16} (53 - 55)^2 = 0.25$$

$$S_{C_1 C_2} = \frac{1}{8} (17 - 34)^2 = 36.1$$

$$S_{C_2 C_3} = \frac{1}{8} (31 - 26)^2 = 3.1$$

$$S_{R_1 R_2} = \frac{1}{8} (39 - 35)^2 = 2.0$$

$$S_e = \frac{1}{8} (36 - 38)^2 = 0.5$$

$$S_f = \frac{1}{8} (17 - 57)^2 = 200.0$$

$$S_G = \frac{1}{8} (38 - 36)^2 = 0.5$$

由上述数据得出方差分析表(如表 46.2-51 所示)。

在这种情况下,一次误差方差  $\sigma_1^2$  与二次误差方差  $\sigma_2^2$  都有  $(18-12)=6$  个过剩份额。一次误差方差平均地分配给  $m$ 、闲、 $A$ 、 $B$  各  $1\sigma_1^2$ ; 二次误差平均分配给  $m$ 、闲、 $A$ 、 $B$ 、 $R'$ 、 $C_1 C_2$ 、 $C_2 C_3$ 、 $D$  各  $0.5\sigma_2^2$ 。因素效应与做 16 次实验一样。求出修正后的误差方差,把闲、 $A$ 、 $C_2 C_3$ 、 $R_1 R_2$ 、 $E$ 、 $G$  合并为一个误差  $e$ ,  $F$  检验就是用  $e$  进行的。

### 6.2.2 显著因素的估计

$$\bar{T} = 60 + \frac{108}{16} = 66.8 \pm 0.7$$

$$\bar{B}_1 = 60 + \frac{37}{8} = 64.6 \pm 1.0$$

$$\bar{B}_2 = 60 + \frac{71}{8} = 68.9 \pm 1.0$$

$$\bar{C}_1 = 60 + \frac{17}{4} - 0.8 = 63.45 \pm 1.4$$

$$\bar{C}_2 = 60 + \frac{65}{8} - 0.8 = 68.1 \pm 1.4$$

$$\bar{C}_3 = 60 + \frac{26}{4} - 0.8 = 67.3 \pm 1.4$$

$$\bar{F}_1 = 60 + \frac{17}{4} = 64.25 \pm 1.2$$

$$\bar{F}_2 = 60 + \frac{57}{4} = 74.25 \pm 1.2$$

计算置信界限时,  $T$ 、 $B$ 、 $C$  的有效反复数取为:

$$16 \times \frac{2}{3}, 8 \times \frac{2}{3}, 4 \times \frac{2}{3}$$

$F$  的有效反复数取为 4。

表 46.2-51 方差分析表

	$J$	$S$	$V$	$E(V)$	修正系数	$V'$	
闲	1	2.2	2.2	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	1.5	
A	1	0.0	0.0	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	0.0	
B	1	72.2	72.2	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	48.1	
$e_1$	0	—	—	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	—	
$R'$	1	100.0	100.0	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	66.7	
C	$C_1 C_2$	1	36.1	36.1	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	24.1
	$C_2 C_3$	1	3.1	3.1	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2 + 8\sigma_m^2$	$\frac{2}{3}$	2.1
D	1	0.2	0.2	$\frac{3}{2}\sigma_2^2$   交错因素	$\frac{2}{3}$	0.1	
$R_1 R_2$	1	2.0	2.0	$\frac{3}{2}\sigma_2^2$   交错因素	$\frac{2}{3}$	2.0	
E	1	0.5	0.5	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2$	$\frac{2}{3}$	0.5	
F	1	200.0	200.0	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2$	$\frac{2}{3}$	200.0	
G	1	0.5	0.5	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2$	$\frac{2}{3}$	0.5	
$e_2$	0	—	—	$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2$	$\frac{2}{3}$	—	
(e)	(17)	(6.7)		$\frac{3}{2}\sigma_2^2 + 4\sigma_1^2$	$\frac{2}{3}$	(0.95)	
				$\sigma_2^2 + 4\sigma_1^2$	1		
				$\sigma_2^2 + 4\sigma_1^2$	1		
				$\sigma_2^2 + 4\sigma_1^2$	1		
				$\sigma_2^2 + 4\sigma_1^2$	1		
计		11					

### 6.2.3 最适宜条件的决定与工序平均估计

最适宜条件为  $B_2 C_2 F_2$ 。以下估计其工序平均。因为因素  $F$  的  $F_1$  实验 12 次,  $F_2$  做了 4 次,所以作为实验,对于  $R_1$ 、 $R_2$ 、 $R_3$  的平均值  $x$ ,有

$$\begin{aligned} \hat{\mu} &= \bar{T} + (\bar{B}_2 - \bar{T}) + (\bar{C}_2 - \bar{T}) + \frac{12}{16}(\bar{F}_2 - \bar{F}_1) \\ &= 66.8 + (68.9 - 66.8) + (68.1 - 66.8) \\ &\quad + \frac{3}{4}(74.2 - 64.2) = 77.7 \end{aligned}$$

$$x = 77.7 \pm 2.2$$

置信界限是按式求出:



$$\pm \sqrt{F \times V_r \times \left\{ \frac{1}{n_r} + \frac{1}{r} \right\}} = \pm \sqrt{5.59 \times 0.96 \left\{ \frac{9}{16} \times \frac{1}{3} \right\}}$$

$$= \pm 2.2$$

式中  $n_r$  是把各因素的实际有效自由度取为 1.5、1.5、1.5、4.5 时求出的。由于  $r$  是  $R_1$ 、 $R_2$ 、 $R_3$  的平均, 所以等于 3。

求单位数麻烦时, 可以用实际有效自由度之和去除总实验数, 计算有效反复数。

对于未进行部分扩大的因素来说, 把对于各显著因素水平的自由度, 乘以修正系数, 便得到实际的有效自由度。

对于进行部分扩大的因素, 在二水平情况下, 未进行部分扩大的而固定的水平, 实际有效自由度取 0.5; 进行部分扩大作了实验的水平, 实际有效自由度取为

4.5。在二水平系的条件下, 实际有效自由度分别取为 0.7 与 12.7。如还有困难可以近似地取  $n_r \approx 1$ 。

## 7 组合产品的实验设计

当遇到水平数不相同的正交试验, 而没有现成的混合正交表可供使用时, 可以选用水平数较多的正交表, 将水平数较少的因素进行两两搭配, 从中选出几种搭配组合因素, 使之能安排在水平数较多的正交表中进行试验, 称组合法:

- 1) 有两套或两种类型零件的设计
- 2) 多因素产品的实验设计
- 3) 面向分类因子的实验设计
- 4) 当部件不能够拆卸和重新装配时实验设计

## 第 3 章 正交实验案例

### 1 瓷砖实验

#### 1.1 目的

##### 1.1.1 制造工序概要

墙用瓷砖的原料是粘土和各种岩石,首先把岩石粉碎与粘土混合,用球磨机进行湿式细磨,脱水,干燥后再进行粉碎,制成含百分之几水分的粉末,把粉末冲压成型后,装入耐火容器烧盆内,用隧道窑进行素烧;然后在瓷砖表面涂釉后,装入烧盆进行本烧制成最终产品。

##### 1.1.2 实验目的

由于装窑方向取直角断面时,外侧接近窑口,温度变化很快,烧成条件不佳,如果在外侧进行素烧就会出现尺寸、形状、外观等方面的次品从而降低收率。这在着色瓷砖上尤为显著。所以为探求在外侧进行着色瓷砖的素烧,可以获得优质着色瓷砖的原料配方为目的进行实验。

##### 1.1.3 实验规模

用小型球磨机调制少量原料,制成与实际产品一道在隧道窑进行烧制,配制用  $L_{27}$  正交表,配制 27 种原料配方,每个配方制作 7 块瓷砖。

### 1.2 因素与水平

#### 1.2.1 配方的因素与水平

根据原料的配方提出的因素与水平时,有种类和数量两方面的问题。

首先,对于粘土的数量,根据经验提出因素的必要性不是很大,所以一般按照经验固定在 20 上。

其次对于粘土的种类,考虑对象为硅目粘土与木节粘土。前者为优品,但由于价格成本差异,考虑在数量上的配比,选择粘土种类为因素。取为因素  $A$ 。

因素  $A$  取三水平:

$A_1$ : 只用硅目粘土;

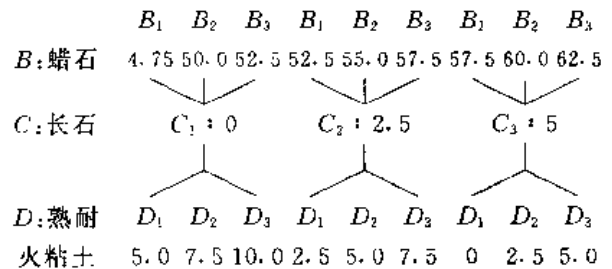
$A_2$ : 只用木节粘土;

$A_3$ : 各用硅目粘土与木节粘土一半。

把两种原料混合起来考虑时的常用方法,可以把

两者的比率取为 8:2.5:5.2:8,但有采购或技术上困难时,也可以采用 7.5:2.5,10:0。

再次把石蜡的数量取为因素  $B$ 。因素  $B$  的水平要与长石数量即因素  $C$  的水平进行组合确定。当蜡石数量增加时长石数量也需要相应的增加。取熟耐火粘土(把不良的产品粉碎混入原料而得)的数量为一个实验因素  $D$ 。从技术上考虑,当蜡石增加时,因素  $D$  应该减少。对于  $B$ 、 $C$ 、 $D$  的水平。以  $C$  为中心,可以进行下列组合:



由于长石的增加与熟耐火粘土的减少正好平衡,所以  $D$  的水平等同于长石+熟耐火粘土之和: 5.0, 7.5, 10.0。当用组合方法决定水平,对消去作用却是有效时,可以采用这种方法。否则最好不要随意采用。

蜡石的种类取为因素  $E$ , 提出三个种类牌号: 三石蜡石, 腾关山蜡石, 谷川蜡石。取表 46.3-1 所示的数字表示各种蜡石对应于蜡石总量的百分比。

表 46.3-1

	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$
三石	100	50	50	50	25	25	0	0	0
腾关山	0	50	25	0	50	25	100	50	50
谷川	0	0	25	50	25	50	0	50	100

还有取长石的种类为因素  $F$ 。此因素与粘土的情况相同,取三水平。

除此作为原料的还有陶石,它等于:

$100 - (20 \text{ 粘土} + B + C + D)$  由于数量上按上式算出,所以无需作为因素提出。在进行配方实验时,某种原料必须作为余数定出,为使合计不超过 100%,通常使用的原料作为余数。实验时,应当把蜡石作为余数的原料,由于需要考虑蜡石的种类,并且与长石进行组合,所以取为因素。剩下的种类陶石只能作为余数的原料。

在上述原料 100% 外,再把某种添加物的添加量取为因素  $G$ 。  $G=0\%$ ,  $G=2.5\%$ ,  $G=5\%$ 。因素  $G$  的数量计入总原料里面可以,如果不超过 10%, 记在 100%

以外也可以。仅为了便于处理添加物。

1.2.2 区组因素、标示因素

因为技术人员对于某一时期进行实验所得的结果,在另一时期是否成立有疑问。主要因为烧成条件的不稳定性,烧成条件是基于经验的。进行反复实验是最理想的,实验规模受限制。

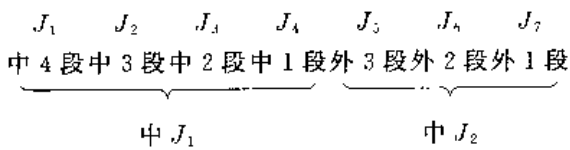
实验目的瓷砖是在外侧素烧,为了比较,也希望在中侧素烧。隧道窑内的烧成条件有上下差异,为了观察这种差异把加热位置取为因素,利用各实验号所得的数块瓷砖进行实验。

要满足两方面的要求,需要大量的瓷砖,所以把数块瓷砖转让给加热位置,作为交换条件,把 27 种配方分三次按一星期左右的时间间隔,通过隧道窑。烧成顺序取为因素  $H$ ,其水平如下:

- $H_1$ :第一次
- $H_2$ :第二次
- $H_3$ :第三次

因素  $H$  与配方因素构成一次因素。包含在正交表  $L_{27}$  的配列中。

加热位置取为因素  $J$ ,其水平如下:



段是指包含在第几段的烧盆,是表现隧道窑的上下的差异。根据配方各制七块瓷砖素烧时,中侧 4 段,外侧 3 段共七个烧盆,各装一块瓷块。因素  $I$  是二次因素。

由于正交表配列有富余。罐磨机不能同时大量细磨 27 种配方所需要的原料,9 台罐磨机要求各使用三次还有停电及其他难以预测的事故,每次细磨条件都存在差异等原因,取罐磨机的顺序为因素  $I$ ,水平为:

- $I_1$ :第一次
- $I_2$ :第二次
- $I_3$ :第三次

同时使用 9 台罐磨机,分配哪一种配方是随机的。

与配方有关的因素  $A, B, C, \dots, G$  是控制因素,  $J$  是标示因素,  $H$  与  $I$  是区组因素。因素  $J$  的效应不是处理对象,只关心  $A \sim G$  的交互作用。因素  $H, I$  即使存在主效应,也无作用,目的仅为了提高实验精度。因素  $H$  在结果上有普遍意义所以有一定意义。

1.2.3 水平对应随机化

应注意对于  $B, C, D$  类的组合选定水平的因素,不要把相互关系弄错。

1.3 配列

1.3.1 一次因素的配列

一次因素的配列使用  $L_{27}(3^{13})$  正交表,配列如下:

因素	$A B C D$	$E$	$F G H I e$
列	2 6 5 8	1 11 12 13	7 9 3 10 4

把表 46.3-2 的水平同正交表的数字对应起来,做成了表 46.3-3 所示的配列表。

根据  $C$  的水平,应注意  $B$  和  $D$  即使水平相同,具体数值也不同。陶石的数量不是因素,但最好计入正交表中。

$C$ :表示长石的数量为 0%。 $C$  与  $G$  存在交互作用,如用拟因素法,可选  $G$  与  $C$  为  $t$  水平,配列于正交表中。

表 46.3-2 因素水平表

一次因素												
$A$ :粘土的种类(数量固定等于 20%)												
$A_1$ :只用木节粘土												
$A_2$ :硅日粘土与木节粘土各半												
$A_3$ :只用硅日粘土												
$B$ :蜡石数量:												
	$B_1$	$B_2$	$B_3$	$B_1$	$B_2$	$B_3$	$B_1$	$B_2$	$B_3$			
	50.0	52.5	57.5	60.0	57.5	62.5	50.0	47.5	52.5			
	↓			↓			↓					
	$C_1$			$C_2$			$C_3$					
	2.5			5			0					
	↓			↓			↓					
	$D_1$	$D_2$	$D_3$	$D_1$	$D_2$	$D_3$	$D_1$	$D_2$	$D_3$			
$C$ :长石数量:												
$D$ :熟耐火粘土:	5	7.5	2.5	2.5	5	0	7.5	10	5			
$E$ :蜡石种类:												
	$E_1$	$E_2$	$E_3$	$E_1$	$E_2$	$E_3$	$E_1$	$E_2$	$E_3$			
三石	0	0	25	50	0	50	100	25	50			
腾光山	50	0	50	50	100	0	0	25	25			
谷川	50	100	25	0	0	50	0	50	25			
$F$ :长石种类												
$F_1$ :三云长石和御花山长石各半												
$F_2$ :只用御花山长石												
$F_3$ :只用三云长石												
$G$ :某种添加物的数量												
	$G_1=0$			$G_2=2.5$			$G_3=5$					
	$(100-20)-(B+C+D)=\text{陶石数量}$											
$H$ :烧成顺序												
$H_1$ :第二次												
$H_2$ :第一次												
$H_3$ :第三次												
$I$ :罐磨机顺序												
$I_1$ :第一次												
$I_2$ :第二次												
$I_3$ :第三次												
二次因素												
$J$ :七个水平												

1.3.2 配列结果

配列结果见表 46.3-3。

表 46.3-3 配列表

No.	因素	(1)(2)(3)(4)(5)(6) (7)(8)(9)(10)(11)										A		B,E			C	F	D		G	H	I
		E	A	H	e	C	B	F	D	G	I	硅 粘上	木 粘土	三 石 蜡	腾 关 山 蜡	谷 川 蜡	三 云 长	御 花 山 长	熟 耐 火 粘	陶 石	添 加 物	烧 成 顺 序	磨 机 顺 序
1	1	1	1	1	1	1	1	1	1	1	-	20	-	27.50	27.50	1.25	1.25	5.0	17.5	-	2	1	
2	2	1	1	1	2	2	2	2	2	2	-	20	-	-	57.50	-	6.00	5.0	12.5	2.5	2	2	
3	3	1	1	1	3	3	3	3	3	3	-	20	13.00	26.50	13.00	-	-	5.0	22.5	5.0	2	3	
4	3	2	2	2	1	1	1	2	2	2	10	10	13.75	27.50	17.50	1.25	1.25	7.5	15.0	2.5	1	2	
5	1	2	2	2	2	2	2	3	3	3	10	10	-	28.75	28.75	-	5.0	17.5	5.0	1	3		
6	2	2	2	2	3	3	3	1	1	1	10	10	-	-	52.50	-	-	7.5	20.0	-	1	1	
7	2	3	3	3	1	1	1	3	3	3	20	-	-	-	55.00	1.25	1.25	2.5	20.0	5.0	3	3	
8	3	3	3	3	2	2	2	1	1	1	20	-	14.25	19.00	14.25	-	5.00	2.5	15.0	-	3	1	
9	1	3	3	3	3	3	3	2	2	2	20	-	-	26.25	26.25	-	-	10.0	17.5	2.5	3	2	
10	4	1	2	3	1	2	3	1	2	3	-	20	26.25	26.25	-	2.50	-	5.0	20.0	2.5	1	3	
11	5	1	2	3	2	3	2	2	3	1	-	20	-	62.50	-	2.50	2.50	5.0	7.5	5.0	1	1	
12	6	1	2	3	3	1	1	3	1	2	-	20	25.00	-	25.00	-	-	5.0	25.0	-	1	2	
13	6	2	3	1	1	2	3	2	3	1	10	10	26.25	-	26.25	2.50	-	7.5	17.5	5.0	3	1	
14	4	2	3	1	2	3	1	3	1	2	10	10	31.25	31.25	-	2.50	2.50	-	12.5	-	3	2	
15	5	2	3	1	3	1	2	1	2	3	10	10	-	50.00	-	-	-	7.5	22.5	2.5	3	3	
16	5	3	1	2	1	2	3	3	1	2	20	-	-	52.50	-	2.50	-	2.5	22.5	-	2	2	
17	6	3	1	2	2	3	1	1	2	3	20	-	31.25	-	31.25	2.50	2.50	2.5	10.0	2.5	2	3	
18	4	3	1	2	3	1	2	2	3	1	20	-	25.00	25.00	-	-	-	10.0	20.0	5.0	2	1	
19	7	1	3	2	1	3	2	1	3	2	-	20	57.50	-	-	-	2.50	5.0	15.0	5.0	3	2	
20	8	1	3	2	2	1	3	2	1	3	-	20	15.00	15.00	30.00	5.00	-	5.0	10.0	-	3	3	
21	9	1	3	2	3	2	1	3	2	1	-	20	24.00	11.75	11.75	-	-	5.0	27.5	2.5	3	1	
22	9	2	1	3	1	3	2	2	1	3	10	10	29.00	14.25	14.25	-	2.50	7.5	12.5	-	2	3	
23	7	2	1	3	2	1	3	3	2	1	10	10	60.00	-	-	5.00	-	-	15.0	2.5	2	1	
24	8	2	1	3	3	2	1	1	3	2	10	10	11.75	11.75	24.00	-	-	7.5	25.0	5.0	2	2	
25	8	3	2	1	1	3	2	3	2	1	20	-	14.25	14.25	29.00	-	2.50	2.5	17.5	2.5	1	1	
26	9	3	2	1	2	1	3	1	3	2	20	-	30.00	15.00	15.00	5.00	-	2.5	12.5	5.0	1	2	
27	7	3	2	1	3	2	1	2	1	3	20	-	47.50	-	-	-	-	10.0	22.5	-	1	3	

3段的瓷砖3块,计算尺寸的标准偏差或者级差R。

1.4 数据分析

1.4.1 特性值

要测量下述特性值:

1) 尺寸:用四边的平均值表示。

2) 不正度:瓷砖偏离正方形的程度。用正方形四边中的最大与最小之差R来表示,差值越小越好。

3) 翘度:瓷砖表面的对角线偏离中心的程度。就两条对角线取平均,表面为凸取正值,凹为负值。一般正号为好。

4) 表面缺陷:某种外观缺点。分为6级 -0、1、2、3、4、5。0级为无缺陷,随着缺陷程度依次升到5级。

5) 龟裂:把瓷砖装入高压罐中,在某种压力下的过热蒸汽中,蒸煮一定时间,观测其上釉表面是否有裂痕。对于经时变化的裂痕进行加速实验观测,变更时间进行反复实验,根据龟裂发生的时间分为1、2、3三级。1级优,3级表面低劣。

6) 尺寸波动:就1)尺寸,取中4级的瓷砖4块、外

1.4.2 表面缺陷数据分析

缺陷程度已分为0~5六个级别。按此六级评分,用计量值进行处理分析。按计量值进行处理,得出大概结论,所得数据如表46.3-4所示。再对每一因素水平求和,做辅助表。方差分析结果如表46.3-5所示。

表 46.3-4 表面缺陷的数据

No.	J <sub>1</sub>	J <sub>2</sub>	J <sub>3</sub>	J <sub>4</sub>	J <sub>5</sub>	J <sub>6</sub>	J <sub>7</sub>
1	0	0	0	0	5	4	5
2	4	4	3	4	5	5	5
3	0	0	0	0	0	0	0
4	0	0	0	0	0	5	5
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	5
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	3	3	5
10	0	0	0	0	0	4	4
11	0	0	0	0	0	0	3
12	0	0	0	0	4	5	5
13	0	0	0	0	0	3	0
14	0	0	0	0	0	2	0
15	0	0	0	0	0	4	4

(续)

No.	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$
16	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0
19	0	0	0	0	4	5	0
20	0	0	0	0	2	5	5
21	0	0	0	0	3	5	5
22	0	0	0	0	0	0	0
23	0	0	0	0	2	2	3
24	0	0	0	0	0	0	0
25	0	0	0	0	0	0	2
26	0	0	2	0	0	0	0
27	0	0	0	0	0	0	0

表 46.3-5 缺陷的方差分析表

因素	$f$	$S$	$V$	$\rho(\%)$
A	2	64.22	32.11	12.2
B	2	12.80	6.40	2.2
C	2	0.38	0.19	—
D	2	6.22	3.11	0.9
E	8	32.38	4.05	5.1
F	2	2.89	1.45	—
G	2	36.22	18.11	6.6
H	2	2.32	1.16	—
I	2	16.67	8.34	2.9
$e_1$	2	2.00	1.00	—
J	6	132.05	22.01	24.5
$A \times J$	12	49.41	4.12	7.8
$B \times J$	12	27.50	2.29	3.6
$C \times J$	12	10.59	0.88	—
$D \times J$	12	1.64	0.14	—
$E \times J$	48	41.84	0.86	—
$F \times J$	12	6.30	0.52	—
$G \times J$	12	30.08	2.51	4.1
$H \times J$	12	30.20	2.52	4.1
$I \times J$	12	11.63	0.97	—
$e_2$	12	8.30	0.69	—
(e)	(116)	(87.89)	(0.76)	26.0
计	188	525.29		100

一次误差  $e_1$  相对于二次误差  $e_2$ , 无显著差异, 而且在一次因素中, 相对于二次因素有很多因素无显著差异, 所以进行并项, 以 (e) 误差检验全部因素。结论是: 控制因素 A、B、D、E、G 有显著差异; 加热位置 I 有显著差异; 交互作用  $A \times J$ 、 $B \times J$ 、 $G \times J$  有显著差异; 主效应 I 有显著差异, 但是区组因素不作考虑;  $H \times J$  有显著差异, 它表示加热位置的温度条件欠佳, 也不作为处理对象, 也排除的。

因此, 首先把 J 分解为  $J_1$  内 ( $J_1 \sim J_4$  的中侧 4 段之间的上下差异)、 $J_2$  内 ( $J_5 \sim J_7$  的外侧 3 段间的上下差异) 和  $J'$  ( $J_1$  与  $J_2$  即中侧与外侧之间的差异) 来观测, 得数据表 46.3-6:

表 46.3-6

因素	$f$	$S$	$V$	$\rho(\%)$
$J_1$ 内	3	0.02	0.01	—
$J_2$ 内	2	19.58	9.79	3.4
$J'$	1	112.45	112.45	21.3
e	116	87.89	0.76	—

由上表可知, 中侧四段的上下差异不显著, 外侧三段的上下差异显著, 中侧与外侧之间的差异显著。误差  $e_1$  与  $e_2$  合并成 e。

其次, 把差异显著的交互作用进行同样分解, 则得如表 46.3-7 所示的数据。其中任一因素与  $J'$  之间的交互作用都有显著差异。

表 46.3-7

因素	$f$	$S$	$V$	$\rho(\%)$
$A \times J_1$ 内	6	0.40	0.07	—
$A \times J_2$ 内	4	6.42	1.61	—
$A \times J'$	2	42.59	21.30	7.8
$B \times J_1$ 内	6	0.46	0.07	—
$B \times J_2$ 内	4	6.49	1.62	—
$B \times J'$	2	20.61	10.31	3.7
$G \times J_1$ 内	6	0.40	0.07	—
$G \times J_2$ 内	4	8.19	2.05	—
$G \times J'$	2	21.49	10.79	3.8
e	116	87.89	0.76	—

再次考虑这些因素, 编制必要的辅助表和估计值的表(如表 46.3-8 所示)。

表 46.3-8 显著因素的估计

	$J_1'$	$J_2'$	计		$A_1$	$A_2$	$A_3$
$A_1$	15	88	103				
$A_2$	0	35	35	$J_1'$	0.42	0.00	0.06
$A_3$	2	16	18	$J_2'$	3.26	1.30	0.59
$B_1$	2	69	71				
$B_2$	15	39	54				
$B_3$	0	31	31				
					$B_1$	$B_2$	$B_3$
$D_1$	2	44	46				
$D_2$	15	53	68				
$D_3$	0	42	42	$J_1'$	0.06	0.42	0.00
				$J_2'$	2.56	1.44	1.44
$G_1$	0	47	47				
$G_2$	15	73	88				
$G_3$	2	19	21				
					$G_1$	$G_2$	$G_3$
$E_1$	0	24	24				
$E_2$	15	24	39	$J_1'$	0.00	0.42	0.06
$E_3$	0	10	10	$J_2'$	1.74	2.70	0.70
$E_4$	0	10	10				
$E_5$	0	11	11				
$E_6$	0	17	17				
					$D_1$	$D_2$	$D_3$
$E_7$	0	16	16		0.73	1.08	0.67
$E_8$	0	14	14				
$E_9$	2	13	15				
计	17	139	156				

$F(0.05) = 3.93$ , 则各个点的置信界限为

$$\pm \sqrt{\frac{3.93 \times 0.76}{n_e}}$$

1.4.3 翘度分析(见表 46.3-9)

表 46.3-9 翘度数据 (mm)

No.	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$
1	0.45	0.45	0.53	0.50	0.95	0.90	1.10
2	0.54	0.55	0.70	0.68	0.93	0.80	0.90
3	0.28	0.30	0.40	0.43	0.80	0.53	0.70
4	0.15	0.23	0.20	0.28	0.75	0.95	0.80
5	0.15	0.30	0.43	0.28	0.28	0.48	0.40
6	0.28	0.33	0.35	0.38	0.95	0.90	0.83
7	0.40	0.60	0.43	0.50	0.63	0.60	0.55
8	0.03	-0.03	-0.08	0.03	0.10	0.18	0.20
9	0.30	0.23	0.43	0.35	0.73	0.63	0.83
10	0.23	0.33	0.33	0.30	0.73	0.90	0.90
11	0.15	0.23	0.20	0.33	0.63	0.58	0.75
12	0.12	0.28	0.40	0.25	1.15	0.88	0.90
13	0.22	0.40	0.45	0.13	0.55	0.40	0.38
14	0.00	0.03	-0.10	0.05	0.50	0.88	0.65
15	0.43	0.23	0.20	0.20	0.65	0.68	0.95
16	0.10	0.03	0.20	0.20	0.25	0.05	0.48
17	0.38	0.15	0.28	0.28	0.80	0.60	0.80
18	0.28	0.25	0.18	0.25	0.38	0.20	0.60
19	0.13	0.33	0.35	0.18	0.60	0.60	0.63
20	0.65	0.55	0.45	0.38	0.80	1.20	1.30
21	0.43	0.35	0.40	0.50	0.88	1.20	0.85
22	0.28	0.25	0.25	0.40	0.83	0.63	0.93
23	0.28	0.30	0.23	0.23	0.38	1.00	0.90
24	0.50	0.30	0.23	0.35	0.38	0.33	0.70
25	0.05	0.30	0.25	0.20	0.48	0.65	0.70
26	0.05	0.13	0.13	0.18	0.40	0.28	0.43
27	0.05	0.15	0.03	0.03	0.48	0.58	0.08

1.4.4 分析结果(见表 46.3-10)

表 46.3-10 翘度的方差分析表

因素	$f$	$S$	$V$	$\rho(\%)$
A	2	2.2052	1.1026	1.7
B	2	0.4464	0.2232	2.1
C	2	0.0675	0.0338	—
D	2	0.1081	0.0540	—
E	8	1.4507	0.1813	6.4
F	2	0.0738	0.0369	—
G	2	0.5741	0.2871	2.8
H	2	0.1175	0.0588	—
I	2	0.1452	0.0726	—
$e_1$	2	0.0335	0.0167	—
$(e_1)$	(12)	(0.5455)	(0.0454)	6.9(一次)
J	6	7.450	1.2417	40.8
A×J	12	0.3509	0.0292	1.7
B×J	12	0.3251	0.0271	1.6
C×J	12	0.1959	0.0163	1.0
D×J	12	0.1123	0.0094	0.6
E×J	48	0.7454	0.0155	3.5
F×J	12	0.1632	0.0136	3.5
G×J	12	0.8569	0.0714	4.4
H×J	12	0.4298	0.0358	2.2
I×J	12	0.0756	0.0063	0.3
$e_2$	12	0.0259	0.0022	13.2(二次)
计	188	18.2358		100.0

就一次因素来说,把V值较小的因素C、D、F、H、I并入一次误差形成( $e_1$ )。用( $e_1$ )进行检验,A、B、E、G有显著差异。

对二次因素来说,其二次误差 $e_2$ 非常小,所以原样进行检验,全部因素都是由显著差异的。

再进行下一步检验,首先分解J进行观测,得表46.3-11中的数据。

表 46.3-11

因素	$f$	$S$	$V$	$\rho(\%)$
$J_1$ 内	3	0.035365	0.011788	0.2
$J_2$ 内	2	0.164713	0.082357	0.9
$J'$	1	7.249972	7.249972	39.7

由上表可知, $J'$ 的影响占绝对优势,从而在交互作用中,把A×J、B×J、G×J加以分解,进行观测,得表46.3-12及46.3-13辅助表。

表 46.3-12

因素	$f$	$S$	$V$	$\rho(\%)$
A× $J_1'$ 内	6	0.025665	0.004278	—
A× $J_2'$ 内	4	0.082450	0.020612	0.4
A× $J'$	2	0.242762	0.121381	1.3
B× $J_1'$ 内	6	0.046594	0.007766	—
B× $J_2'$ 内	4	0.036257	0.009064	—
B× $J'$	2	0.242217	0.121108	1.3
E× $J_1'$ 内	24	0.128377	0.005349	—
E× $J_2'$ 内	16	0.428309	0.026769	2.3
E× $J'$	8	0.188721	0.023590	1.0
G× $J_1'$ 内	6	0.044838	0.007473	—
G× $J_2'$ 内	4	0.086794	0.021699	0.5
G× $J'$	2	0.725280	0.362640	4.0

表 46.3-13 辅助表

	$J_1'$	$J_2'$	计		$J_1'$	$J_2'$	计
$A_1$	13.71	23.11	36.82	$E_1$	4.40	6.30	10.70
$A_2$	9.24	18.06	27.30	$E_2$	5.82	7.09	12.91
$A_3$	6.92	13.29	20.21	$E_3$	2.22	5.01	7.23
				$E_4$	2.04	5.74	7.78
$B_1$	11.55	20.31	31.36	$E_5$	2.50	5.04	7.54
$B_2$	9.37	14.99	24.36	$E_6$	3.34	6.46	9.80
$B_3$	8.95	19.16	28.11	$E_7$	1.89	5.85	7.74
				$E_8$	4.51	6.54	11.05
$G_1$	7.99	19.28	27.27	$E_9$	3.15	6.43	9.58
$G_2$	11.35	21.37	32.72				
$G_3$	10.53	13.81	24.34	计	29.87	54.46	84.33

根据上表,二次因素只考虑了 $J_1'$ 、A× $J'$ 、B× $J'$ 、G× $J'$ 。

然后求平均值,得出表46.3-14。

表 46.3-14

	$A_1$	$A_2$	$A_3$	$B_1$	$B_2$	$B_3$	$G_1$	$G_2$	$G_3$
$J_1'$	0.38	0.26	0.19	0.32	0.26	0.25	0.22	0.32	0.29
$J_2'$	0.86	0.67	0.49	0.75	0.56	0.71	0.71	0.79	0.51

就各个特性的分析结果可以画成曲线图进行分析,进行综合比较研究。加大规模,提高精度,参照日常生产的检验标准,通过综合各个特性的收率,来推断结果。

1.4.5 因素与水平

规定如下:首先把显著影响各个特性的添加物数量,选为最重要因素。至于水平取 $A_1=50\%$ (改进), $A_2=1\%$ (现行)。其次把不同牌号添加物粒度为因素。

B: 添加物粒度

$B_1$  = 粗粒度 (现行)

$B_2$  = 细粒度 (试验)

C: 蜡石数量

$C_1$  = 43% (试验)

$C_2$  = 53% (现行)

D: 蜡石种类

$D_1$  = 现行配方

$D_2$  = 试验配方

因为原料的装入量影响罐磨机的细磨程度, 所以取装入量为因素。

E: 装入量

$E_1$  = 1300kg

$E_2$  = 1200kg (现行)

F: 熟耐火粘土数量

$F_1$  = 0%

$F_2$  = 1%

G: 长石数量

$G_1$  = 0%

$G_2$  = 5%

1.4.6 配列

正交配列使用  $L_{16}(2^{15})$  正交表。各因素配列如下:

其中 R 是表现第一次与第二次的区组因素。

因素 R A B C D E F G

列 (1) (3) (4) (6) (8) (10) (12) (14)

如若只考虑  $R_1$ , 因素 A~G 可构成  $L_8(2^7)$  的正交表, 首先只考虑  $R_1$  的实验, 进行数据分析, 就能知道 A~G 的主效应的相对重要程度。如果认为实验还不充分, 就要追加  $R_2$  的实验, 以提高精度并且提出 A 与 B~G 等因素间的交互作用。

1.4.7 数据与分析

各个配方约制造 1000 块瓷砖, 各分 500 块, 在中侧 ( $K_1$ ) 与外侧 ( $K_2$ ) 进行素烧, 结果如表 46.3-15 所示。

表 46.3-15 配列与数据

No.	A	B	C	D	E	F	G	$K_1$	$K_2$	计
1	1	1	1	1	1	1	1	83.5	52.7	136.2
2	1	1	1	2	2	2	2	82.8	59.3	142.1
3	1	2	2	1	1	2	2	87.9	80.5	168.4
4	1	2	2	2	2	1	1	94.4	71.6	166.0
5	2	1	2	1	2	1	2	94.0	47.6	141.6
6	2	1	2	2	1	2	1	32.5	0.0	32.5
7	2	2	1	1	2	2	1	58.1	19.5	77.6
8	2	2	1	2	1	1	2	74.3	0.0	74.3
计								607.5	331.2	938.7

作辅助表 (表 46.3-16) 进行方差分析 (表 46.3-17)。

表 46.3-16 辅助表

	$K_1$	$K_2$	计
$A_1$	348.6	264.1	612.7
$A_2$	258.9	67.1	326.0
$B_1$	292.8	159.6	452.4
$B_2$	314.7	171.6	486.3
$C_1$	298.7	131.5	430.2
$C_2$	308.8	199.7	508.5
$D_1$	323.5	200.3	523.8
$D_2$	284.0	130.9	414.9
$E_1$	278.2	133.2	411.4
$E_2$	329.3	198.0	527.3
$F_1$	346.2	171.9	518.1
$F_2$	261.3	159.3	420.6
$G_1$	268.5	143.8	412.3
$G_2$	339.0	187.4	526.4
计	607.5	311.2	938.7

表 46.3-17 方差分析表

因素	f	S	V
A	1	5137.31	
B	1	71.83	○
C	1	383.18	
D	1	741.20	
E	1	839.55	
F	1	594.14	
G	1	813.68	
K	1	4771.36	
A×K	1	719.58	
B×K	1	6.13	○
C×K	1	210.98	△
D×K	1	55.88	○
E×K	1	11.73	○
F×K	1	326.71	
G×K	1	45.23	○
(e)	(5)	(190.80)	(36.16)
计	15	14728.49	

由于小型实验, 可以把小的因素集合起来形成误差, 就是指  $B \times K, D \times K, E \times K, G \times K$ 。主效应中只有 B 无显著差异, 把 B 以上四个交互作用考虑为误差, 进行估计确定最宜条件。误差方差  $V_e = 38.16 (f=5)$ , 求主效应置信界限为  $\pm 0.5$ , 在二元表中的置信界限为  $\pm 7.1$  (如表 46.3-18 所示)。

表 46.3-18 效应估计

	$A_1$	$A_2$	$C_1$	$C_2$	$F_1$	$F_2$
$K_1$	87.2	64.7	74.7	77.2	86.5	65.3
$K_2$	66.0	16.8	32.9	49.9	43.0	39.8
					$\pm 7.1$	
$(B_1: 56.6)$	$D_1: 65.0$		$E_1: 51.4$		$G_1: 51.5$	
$(B_2: 60.8)$	$D_2: 51.9$		$E_2: 65.9$		$G_2: 65.8$	
					$\pm 5.0$	

描述曲线如图 46.3-1 所示。

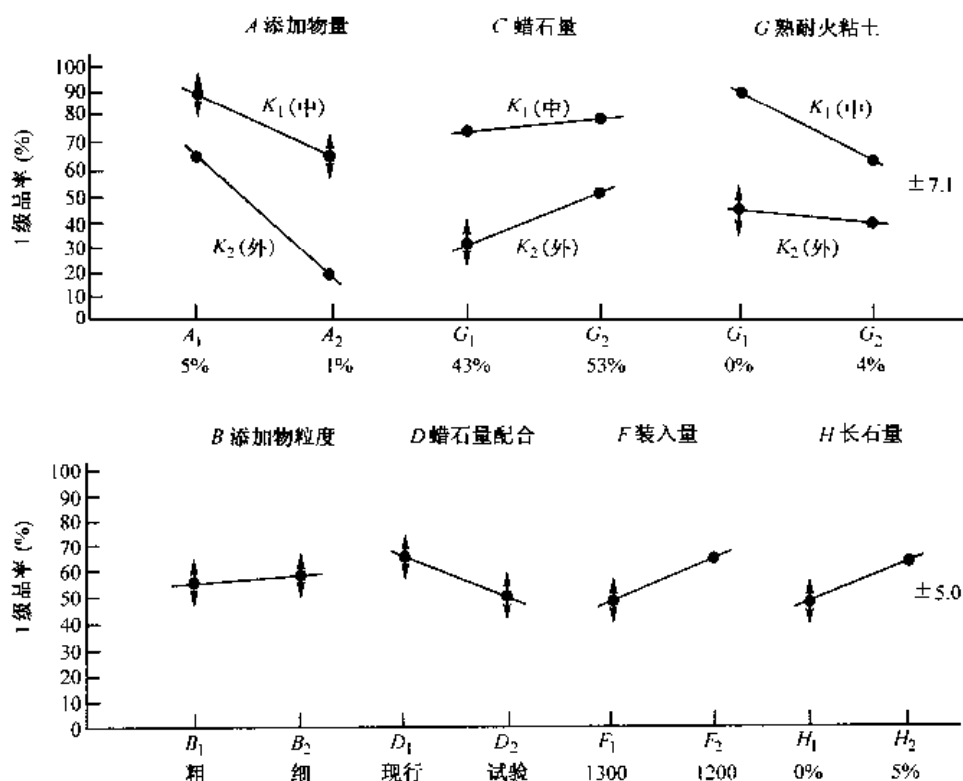


图 46.3-1 收率的曲线

1.4.8 估计与处理

$$n_e = \frac{16}{11} - \pm 8.6(\%)$$

根据曲线图最宜条件为 A<sub>1</sub>C<sub>2</sub>D<sub>1</sub>E<sub>2</sub>F<sub>1</sub>G<sub>2</sub> 的组合, 按外侧 K 估计工序平均。

如果算术加法的话, 公式为

$$\hat{\mu} = \overline{A_1 K_2} + \overline{C_2 K_2} + \overline{F_1 K_2} + \overline{D_1} + \overline{E_2} + \overline{G_2} - 2 \overline{K_2} - 3 \overline{T}$$

也可以用欧米加法进行估计, 把上式改为分贝值。

由于  $\overline{K_2} = 41.4\%$ ,  $\overline{T} = 58.7\%$ , 所以

$\hat{\mu}$  的分贝值 = 8.75(dB)

由此得

$$\hat{\mu} = 88.2 \pm 8.6(\%) \text{ 式中 } \pm 8.6\% \text{ 是置信界限}$$

$$\pm \sqrt{F \times V \times \hat{\mu}(1-\hat{\mu}) \times \frac{1}{T(1-T)}} \times \frac{1}{n_e} \text{ 其中}$$

2 电线的 PVC 绝缘材料实验

2.1 目的

本实验的目的是找出对聚氯乙烯 PVC 电线的脆化特性及其他质量特性有影响的最宜可塑剂、稳定剂、填充剂的配和方法 and 加工条件。

2.2 因素与水平

规定聚氯乙烯树脂为 100 份, 可塑剂 (DOP) 的最低需要为 30 份, 滑剂一份, 滚筒的辊炼条件予以固定, 选取因素与水平如表 46.3-19 所示。

表 46.3-19 因素与水平

配合剂重量/份							加工条件	
可塑剂		稳定剂		填充剂			挤压机	
A	B	C	D	E	F	G	K	L
DOA	n-DOP	三监基 硫酸铅	三监基 磷酸铅	生纤维 材料	氧化钛	碳	螺旋的转数 /r · min <sup>-1</sup>	气缸温度
A <sub>1</sub> =0	B <sub>1</sub> =10	C <sub>1</sub> =0	D <sub>1</sub> =3	E <sub>1</sub> =0	F <sub>1</sub> =0	G <sub>1</sub> =0	K <sub>1</sub> =80	L <sub>1</sub> =170
A <sub>2</sub> =10	B <sub>2</sub> =20	C <sub>2</sub> =2.5	D <sub>2</sub> =5	E <sub>2</sub> =5	F <sub>2</sub> =5	G <sub>2</sub> =25	K <sub>2</sub> =95	L <sub>2</sub> =175
A <sub>3</sub> =20	B <sub>3</sub> =30	C <sub>3</sub> =5	D <sub>3</sub> =7	E <sub>3</sub> =10	F <sub>3</sub> =10	G <sub>3</sub> =3	K <sub>3</sub> =110	L <sub>3</sub> =180



### 2.3 配列与数据

为了尽可能只分析 A、B、C、D、E、F、G、K、L 等 9 个因素的主效应,利用  $L_{27}(3^9)$  正交表。做如表 46.3-20 配列:

总共 27 次实验是按随机顺序进行的,数据为特性测量值。正交表的配列与数据如表 46.3-21,其中括弧内的数值是左边脆化温度测量值加上虚拟平均值 25°C

后的数值。

表 46.3-20 配列表

因素	列	因素	列
A	1	e	8
B	2	e	9
C	3	e	10
D	4	K	11
E	5	L	12
F	6	e	13
G	7		

表 46.3-21 实际配列与实验数据

因素 实验号	配合剂重量(份)							加工条件		实验结果	
	A	B	C	D	E	F	G	K /r/min	L /°C	/°C	
1	0	10	0	3	0	0	0	80	170	+5	(30)
2	0	10	0	3	5	5	2.5	95	175	+2	(27)
3	0	10	0	3	10	10	5	110	180	+8	(33)
4	0	20	2.5	5	0	0	0	110	180	-15	(10)
5	0	20	2.5	5	5	5	2.5	80	170	-6	(19)
6	0	20	2.5	5	10	10	5	95	175	-10	(15)
7	0	30	5	7	0	0	0	95	175	-28	(-3)
8	0	30	5	7	5	5	2.5	110	180	-19	(6)
9	0	30	5	7	10	10	5	80	170	-23	(2)
10	10	10	2.5	7	0	5	3	80	175	-13	(12)
11	10	10	2.5	7	5	10	0	95	180	-17	(8)
12	10	10	2.5	7	10	0	2.5	110	170	-7	(18)
13	10	20	5	3	0	5	5	110	170	-23	(2)
14	10	20	5	3	5	10	0	80	175	-31	(-6)
15	10	20	5	3	10	0	2.5	95	180	-23	(2)
16	10	30	0	5	0	5	5	95	180	-34	(-9)
17	10	30	0	5	5	10	0	110	170	-37	(-12)
18	10	30	0	5	10	0	2.5	80	175	-29	(-4)
19	20	10	5	5	0	10	2.5	80	180	-27	(-2)
20	20	10	5	5	5	0	5	95	170	-27	(-2)
21	20	10	5	5	10	5	0	110	175	-30	(-5)
22	20	20	0	3	0	10	2.5	110	175	-35	(-10)
23	20	20	0	3	5	0	5	80	180	-35	(-10)
24	20	20	0	3	10	5	0	95	170	-38	(-13)
25	20	30	2.5	7	0	10	2.5	95	170	-39	(-14)
26	20	30	2.5	7	5	0	5	110	175	-40	(-15)
27	20	30	2.5	7	10	5	0	80	180	-41	(-16)

### 2.4 数据分析

只考虑主效应问题,所以各因素的水平,例如  $A_1$ , 是  $A_1$  进行九次实验所有数据的合计。 $A_2$ 、 $A_3$  做同样的处理。各因素的水平合计:

$A_1$	$A_2$	$A_3$	计
139	11	87	63
$B_1$	$B_2$	$B_3$	计
119	9	-65	63
$C_1$	$C_2$	$C_3$	计
32	37	-6	63
$D_1$	$D_2$	$D_3$	计
43	10	10	63
$E_1$	$E_2$	$E_3$	计
16	15	32	63
$F_1$	$F_2$	$F_3$	计
26	23	14	63

$G_1$	$G_2$	$G_3$	计
-7	42	28	63
$K_1$	$K_2$	$K_3$	计
23	12	28	63
$L_1$	$L_2$	$L_3$	计
31	12	20	63

各因素的主效应计算(取整处理)

$$\text{修正项} = CF = \frac{(\text{合计})^2}{27} = \frac{63^2}{27} = 147$$

$$S_A = \frac{1}{9} (A_1^2 + A_2^2 + A_3^2) - CF = \frac{1}{9} [139^2 + 11^2 + (-87)^2] - 147 = 2854$$

$$S_B = \frac{1}{9} [119^2 + 9^2 + (-65)^2] - 147 = 1905$$

$$S_C = \frac{1}{9} [32^2 + 37^2 + (-6)^2] - 147 = 122$$

$$S_D = \frac{1}{9} [43^2 + 10^2 + 10^2] - 147 = 80$$

$$S_E = \frac{1}{9} [16^2 + 15^2 + 32^2] - 147 = 20$$

$$S_F = \frac{1}{9} [26^2 + 23^2 + 14^2] - 147 = 8$$

$$S_G = \frac{1}{9} [(-7)^2 + 42^2 + 28^2] - 147 = 142$$

$$S_K = \frac{1}{9} [23^2 + 12^2 + 28^2] - 147 = 5$$

$$S_L = \frac{1}{9} [31^2 + 12^2 - 20^2] - 147 = 20$$

$$S_T = [30^2 + 27^2 + \dots + (-16)^2] - 147 = 5202$$

$$S_e = S_T - (\text{各因素的波动总计}) = 36$$

有时可以把因素波动分解为自由度为1的正交多项式成分,分解如下:

$$S_A = \begin{cases} \text{一次} = \frac{1}{2 \times 9} [139 \times (-1) + (-87)]^2 = 2837 \\ \text{二次} = 2854 - 2837 = 17 \end{cases}$$

$$S_B = \begin{cases} \text{一次} = \frac{1}{2 \times 9} [119 \times (-1) + (-65)]^2 = 1881 \\ \text{二次} = 1905 - 1881 = 24 \end{cases}$$

$$S_C = \begin{cases} \text{一次} = \frac{1}{2 \times 9} [32 \times (-1) + (-6)]^2 = 80 \\ \text{二次} = 122 - 80 = 42 \end{cases}$$

$$S_D = \begin{cases} \text{一次} = \frac{1}{2 \times 9} [43 \times (-1) + 10]^2 = 61 \\ \text{二次} = 80 - 61 = 19 \end{cases}$$

$$S_e = \begin{cases} \text{一次} = \frac{1}{2 \times 9} [(-7) \times (-1) + 28]^2 = 68 \\ \text{二次} = 142 - 68 = 74 \end{cases}$$

方差分析得表 46.3-22。

表 46.3-22 方差分析表

因素	f	S	V	F	f	V	f
A	2	2854	1427	317.2	{一次 1	2837	630
					{二次 1	17	—
B	2	1905	952	211.5	{一次 1	1881	418.8
					{二次 1	24	5.34
C	2	122	61	13.6	{一次 1	80	17.8
					{二次 1	42	9.3
D	2	80	40	8.9	{一次 1	61	13.45
					{二次 1	19	4.22
E	2	20	10	2.2			
F	2	8	4	—			
G	2	142	71	15.8	{一次 1	68	15.1
					{二次 1	74	16.5
K	2	15	8	1.8			
L	2	20	10	2.2			
e	8	36	4.5		(e)15	99	6.18
计	26	5202					

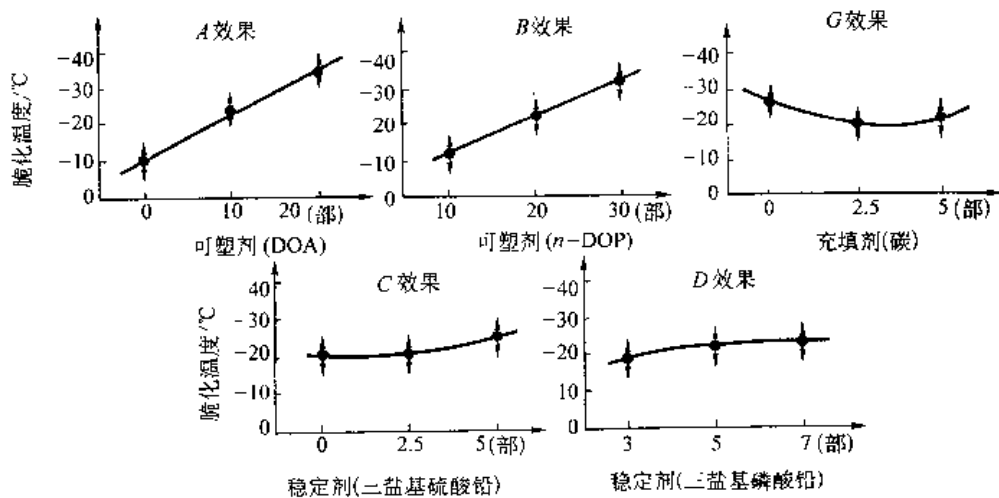


图 46.3-2 主效应曲线图

把各显著效应因素的效应画成曲线图(见图 46.3-2)。

置信界限:

$$\pm \sqrt{F \times V \times \frac{1}{n_e}} = \pm \sqrt{4.49 \times 6.18 \times \frac{1}{9}} = \pm 1.8$$

### 2.5 结论

因素 A、B 原来就是作为提高脆化特性的可塑剂来配用的,可以从 A、B 的效应图看出脆化特性的提高程度。稳定剂 C、D 对脆化特性有一定的影响,但影响程度很小,可不作考虑。填充剂 C 对脆化特性的影响

在选取 2.5 份左右时最坏,有必要增加数量进行实验。取任一配合,例如取  $A_2$ 、 $B_2$ 、 $C_1$ 、 $D_1$ 、 $G_3$  时,用下式求脆化温度的变化。

$$\begin{aligned} \hat{\mu} &= \text{虚拟平均} + \frac{A_2}{9} + \frac{B_2}{9} + \frac{C_1}{9} + \frac{D_1}{9} + \frac{G_3}{9} - 4 (\text{总平均}) \\ &= -25 + \frac{11}{9} + \frac{9}{9} + \frac{32}{9} + \frac{43}{9} + \frac{28}{9} - 4 \times \frac{63}{27} \\ &= -21 \pm 3.8 \end{aligned}$$

在此条件下制造聚氯乙烯电线,其脆化温度的预测值等于:

$n_c = \frac{\text{总实验数}}{\text{上述估计式中所考虑的因素自由度之和}}$

$$= \frac{27}{2 \times 5 + 1} = \frac{27}{11}$$

$$x = \bar{\mu} \pm \sqrt{F \times V_e \times \left\{ \frac{1}{n_c} + 1 \right\}}$$

$$= -21.0 \pm \sqrt{4.49 \times 6.18 \times \left\{ \frac{11}{17} + 1 \right\}}$$

$$= -21.0 \pm 6.8$$

因而以  $A_2B_2C_1D_1G_3$  条件制造聚氯乙烯电线,其脆化温度的每个测量值处于  $-21^\circ\text{C} \pm 6.8^\circ\text{C}$  的范围内。其置信界限对任一组合为常值。对于脆化温度更低的组合  $A_1B_3C_3D_2G_1$  对脆化温度的估计值为  $-51^\circ\text{C} \pm 6.8^\circ\text{C}$ 。

### 2.6 说明

补充说明实验设计的改进问题。

首先,在因素选择与水平决定方面。一般原料调配实验应该把配合物资的质量差异与数量同时加以考虑,在调查 DOP、DOA、n-DOP 三种可塑剂的最宜配合时,通常按下述两种方式选取因素与水平。

1) 在 DOP、DOA、n-DOP 三种可塑剂中,决定选哪一种最好和加入量多少时,不考虑三种混合,而选定两个因素

A: 表现种类因素  $A_1 = \text{DOP}, A_2 = \text{DOA}, A_3 = \text{n-DOP}$

B: 表现数量的因素,对应于 A 水平的取值如表 46.3-23 所示:

表 46.3-23

	$B_1$	$B_2$	$B_3$
$A_1$	30	50	70
$A_2$	24	40	56
$A_3$	30	50	70

对应于  $A_1, B_1$  的取值为最小,  $B_2$  中间值,  $B_3$  最大。

2) 在考虑 DOP、DOA、n-DOP 的混合时,通常取下述三个因素。

A: 可塑剂的合计量

$A_1 = 30$  份,  $A_2 = 50$  份,  $A_3 = 70$  份

B: 可塑剂总量中, DOP 的比例

C: 可塑剂总量中, DOA 的比例

n-DOP 的配合比由 BC 自然可以算出,不设为因素。数值如表 46.3-24 所示:

表 46.3-24

	1	2	3	4	5	6	7	8	9
B(DOP)	100	50	50	50	25	25	0	0	0
C(DOA)	0	50	25	0	50	25	100	50	0
n-NOP 的比例 (不是因素)	0	0	25	50	25	50	0	50	100

其次,本实验只提出了九个因素。最好能再提出一些更好的因素,同时改变水平的困难性未反应在配列中。由

于有空闲列,至少希望进行 A 与 B 的交互作用实验。

另外各组合抽取很多样本,如果每个实验能取三、四个以上的测量值进行观察各种温度下是否脆化。每个实验号取几个测量值作一下波动分析。

特性值间的相互关系,可以把各因素对各种特性值的影响画在同一曲线图中表示或者调查 27 种特性值的互相关系。还有以  $5^\circ\text{C}$  为间隔,收集脆化温度数据,即每个实验号各取 2 个实验样品,按  $5^\circ\text{C}, 0^\circ\text{C}, -5^\circ\text{C}, -10^\circ\text{C}$  进行实验。如果  $-5^\circ\text{C}$  良好,而  $-10^\circ\text{C}$  脆化则用  $-7.5^\circ\text{C}$  为分级值。

## 3 脱蜡装置的灌输实验

### 3.1 目的

根据溶剂精炼法炼制高级润滑油。在溶解抽出工序后,为了炼制凝固点低的优质油,以降低凝固点为目的,研究溶剂脱蜡装置的操作条件的改进问题。脱蜡以后的油即脱蜡油,是高价油,实验日的尽可能地在凝固电的状态下提高收率。

### 3.2 因素与水平

先介绍该装置-巴利札尔脱蜡装置,工序如图 46.3-3 所示。

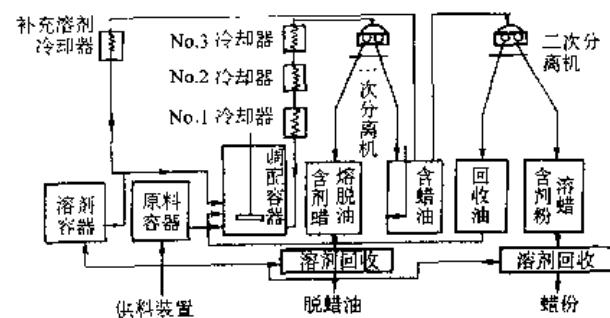


图 46.3-3 巴利札尔脱蜡装置

操作如下:

- 1) 把抽出工序送来的脱蜡原料油和混合剂装入调配油罐中。以一定的温度加热后,进行搅拌溶解。
- 2) 通过 No. 1、No. 2、No. 3 号冷却器冷却到规定温度,送入第一次离心分离器。
- 3) 析出的石蜡不能溶于溶剂,油可以溶于重二氯化乙烯和苯的混合溶剂,进行分离,中心部分残留的石蜡可用喷嘴取出。
- 4) 喷嘴位置能够移动大约 5mm,喷嘴位置决定着脱蜡油与含蜡油的比例,也影响脱蜡油的凝固点。
- 5) 为从含蜡油中进一步炼油,用冷却器冷却补充溶剂之后,加入大量的含蜡油,进行二次离心分离。
- 6) 产出的回收油,再送入调配油罐。

7) 加热并回收脱蜡油中的溶剂和石蜡中的溶剂。  
 操作人员为 5 名, 设备自动作业。由于使用实际的  
 制造工序进行实验, 每个因素都取两水平, 现行水平和  
 改进的新水平。

A: 原料油

A<sub>1</sub> = 重原料油 (V<sub>18</sub>R100°C65)

A<sub>2</sub> = 轻原料油 (V<sub>18</sub>R100°C58)

(重、轻原料油为标示因素, 未选水平)

B: 溶剂中苯的含量

B<sub>1</sub> = 22%, B<sub>2</sub> = 24%

其中增加苯的用量时, 可以降低冷却温度, 所以取  
 24% 时, 把温度降低 1°C 使用。

C: 调配油罐温度

C<sub>1</sub> = 30°C, C<sub>2</sub> = 35°C

D: 原料油与溶剂的调配比

D<sub>1</sub> = 1 : 3.8, D<sub>2</sub> = 1 : 4.6 溶剂越多收率越高。

E: 冷却温度

E<sub>1</sub> = -17°C, E<sub>2</sub> = -15°C

F: 冷却速度

F<sub>1</sub> = 使用 2 台冷却器

F<sub>2</sub> = 使用 3 台冷却器

G: 离心分离机的喷嘴位置

G<sub>1</sub> = 经验确定的位置

G<sub>2</sub> = 比 G<sub>1</sub> 确定的位置靠近中心 1/10

H: 离心分离机使用台数

H<sub>1</sub> = 9 台

H<sub>2</sub> = 8 台

I: 补充溶剂温度

I<sub>1</sub> = -20°C

I<sub>2</sub> = -17°C

J: 补充溶剂的流量

J<sub>1</sub> = 2.8 kL/h

J<sub>2</sub> = 3.3 kL/h

### 3.3 配列与数据

原料油为一次因素, 苯为二次因素, 其他全部为三  
 次因素。正交配列如表 46.3-25 所示。

表 46.3-25 配列表

实验 顺序	因素 No.	A		B			D				E		F		C		H		
		1	2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
①	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	1	2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	
	4	3	1	1	1	1	2	2	2	1	1	1	2	2	2	2	2	2	
	3	4	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1	
②	8	5	1	2	2	2	1	1	2	1	1	2	2	1	1	2	2	2	
	5	6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	1	
	7	7	1	2	2	2	2	1	1	1	2	2	2	2	2	1	1	1	
	6	8	1	2	2	2	2	1	1	2	1	1	1	1	1	2	2	2	
③	12	9	2	1	2	2	1	2	2	1	2	1	2	1	2	1	2	2	
	10	10	2	1	2	1	2	2	2	2	1	2	1	2	1	2	1	1	
	9	11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	1	
	11	12	2	1	2	2	1	2	1	2	1	1	1	2	1	2	1	2	
④	13	13	2	2	1	2	1	2	2	1	2	2	1	1	2	2	2	1	
	15	14	2	2	1	2	2	2	1	2	1	2	2	1	1	1	2	2	
	16	15	2	2	1	2	1	2	1	1	2	2	1	2	1	1	2	2	
	14	16	2	2	1	2	1	2	2	2	1	1	2	1	2	2	2	1	
水平		1	58	22	1 : 3.8 - 20°C			-15°C 3.3 现行 3 台 30°C 9 台											
		2	65	24	1 : 4.6 - 17°C			-17°C 2.8 缩小 1/10 2 台 35°C 8 台											

其中因素的水平与配列表中的数字随机相对应。

(续)

为消除中途的在制品影响, 一次实验进行 24h。收  
 率测量了 4 个点, 凝固点测量了 3 个点。表 46.3-26 中  
 数值为平均值, 其中 No. 7 是欠测值, 用最小误差方法  
 估计。

表 46.3-26 数据

实验号 No.	收率提高 (%)	凝固点 /°C
1	4.8	-17.5
2	3.0	-16.7
3	5.2	-12.5
4	12.2	-16.7

实验号 No.	收率提高 (%)	凝固点 /°C
5	1.8	-15.0
6	3.2	-16.3
7	(7.3)	(-15.5)
8	2.5	-17.5
9	0.8	-12.5
10	1.8	-17.5
11	4.2	-15.0
12	0.7	-16.7
13	-0.6	-14.2
14	-5.3	-18.3
15	4.0	-13.3
16	2.0	-14.2

3.4 数据分析

在表 46.3-26 中,列出了凝固点的平均值,但不是真正的平均值。因为重复间的波动,形成高次误差,由于重复间的误差非常小,也就忽略了。

$L_{16}$ 配列所用的辅助表计算。使用配列型(5)(见表

46.3-27),可以省略 5 次加法运算。配列型(5)有 5 根线段,如果对应于其两端顶点的两列数字组合,能求出合计,则作为  $2 \times 2$  的二元表的交互作用,可以调查线段列的效应。本例中可以用第一条线段对应于第一列与第二列的各个组合求出合计。

由此得收率的  $S$  值如下:

表 46.3-27 用配列型(5)的辅助表

列 1 2 3	收率	凝固点	列 4 8 12	收率	凝固点	列 5 10 15	收率	凝固点
1 1 1	250	-63.4	1 1 1	68	-59.2	1 1 1	142	-63.0
1 2 2	148	-64.3	1 2 2	27	-68.8	1 2 2	95	-61.7
2 1 2	75	-61.7	2 1 2	205	-56.3	2 1 2	30	-60.8
2 2 1	1	-60.0	2 2 1	174	-65.1	2 2 1	207	-63.9
A A B B	474	-249.4	D E D E	474	-249.4	e C H	474	-249.4
58 22			3.8-15°C			现 9		
65 24			4.6-17°C			1/10 8		

列 7 9 4	收率	凝固点	列 6 11 3	收率	凝固点
1 1 1	75	-68.0	1 1 1	131	-65.8
1 2 2	91	-63.4	1 2 2	131	-58.9
2 1 2	106	-59.2	2 1 2	83	-59.7
2 2 1	202	-58.8	2 2 1	129	-65.0
B J C D	474	-249.4	L B F E	474	-249.4
3.3 30°C			-20°C 3		
2.8 35°C			-17°C 2		

$$S_A = \frac{(398-76)^2}{16} = \frac{322^2}{16} = 6480$$

$$S_B = \frac{(250-222)^2}{16} = \frac{28^2}{16} = 49$$

$$S_{A \times B} = \frac{(325-149)^2}{16} = \frac{176^2}{16} = 1936$$

$$S_k = \frac{(273-201)^2}{16} = 324$$

$$S_l = 0$$

$$S_G = \frac{(172-302)^2}{16} = 1056$$

$$S_H = \frac{(349-125)^2}{16} = 3136$$

$$S_{B \times D} = \frac{(166-308)^2}{16} = 1260$$

$$S_J = \frac{(181-293)^2}{16} = 784$$

$$S_C = \frac{(277-197)^2}{16} = 400$$

$$S_I = \frac{(262-212)^2}{16} = 156$$

$$S_{B \times E} = \frac{(214-260)^2}{16} = 132$$

$$S_F = \frac{(260-214)^2}{16} = 132$$

凝固点的  $S$  值如下:

$$S_A = \frac{(-127.7+121.7)^2}{16} = 2.25$$

$$S_B = \frac{(-123.4+126.0)^2}{16} = 0.42$$

$$S_{A \times B} = \frac{(-125.1+124.3)^2}{16} = 0.04$$

$$S_D = \frac{(-128.0+121.4)^2}{16} = 2.72$$

$$S_E = \frac{(-115.5+133.9)^2}{16} = 21.16$$

$$S_{D \times E} = \frac{(-124.3+125.1)^2}{16} = 0.04$$

$$S_e = 0$$

$$S_G = \frac{(-123.8+125.6)^2}{16} = 0.20$$

$$S_H = \frac{(-126.9+122.5)^2}{16} = 1.21$$

$$S_{B \times D} = \frac{(-131.4+118.0)^2}{16} = 11.22$$

$$S_J = \frac{(-127.2+122.2)^2}{16} = 1.56$$

$$S_C = \frac{(-126.8+122.6)^2}{16} = 1.10$$

$$S_I = \frac{(-124.7+124.7)^2}{16} = 0$$

$$S_{B \times E} = \frac{(-125.5+123.9)^2}{16} = 0.16$$

$$S_F = \frac{(-130.8 - 118.6)^2}{16} = 9.30$$

本例有欠测值,误差的自由度为0,所以不能用误差方差进行检验因素。一般来说,在误差自由度小,检验精度不可靠的情况下,可以取效应小的因素自由度大约一半,作为误差进行检验。一般经验而言,把大约一半的因素效应并入误差是符合实际的,并且是危险少的方法,并入误差的因素也可以进行检验。方差分析如表 46.3-28 和表 46.3-29 所示。

表 46.3-28 对应于收率的方差分析表

因素	f	S	F	p(%)
A	1	6480	37.8	30.0
B	1	49○	-	-
A×B	1	1936	11.3	8.4
C	1	400○	2.3	-
D	1	5112	30.0	23.6
B×D	1	1260	7.4	5.2
E	1	324○	1.9	-
B×E	1	132○	-	-
D×E	1	6○	-	-
F	1	132○	-	-
G	1	1056	5.9	4.2
H	1	3136	18.4	14.1
I	1	156○	-	-
J	1	784△	4.6	3.0
(e)	(7)	(1199)	(171)	(11.4)
计	14	20963		100.0

表 46.3-29 对应于凝固点的方差分析表

因素	f	S	F	p(%)
A	1	2.52	9.0	4.3
B	1	0.42○	1.5	-
A×B	1	0.04○	-	-
C	1	1.10○△	3.9	-
D	1	2.72	9.7	4.7
B×D	1	11.22	40.1	21.3
E	1	21.16	75.6	40.5
B×E	1	0.16	-	-
D×E	1	0.04	-	-
F	1	9.30	33.2	17.5
G	1	0.20○	-	-
H	1	1.21△	4.4	-
I	1	0.00	-	-
J	1	1.56	5.6	2.5
(e)	(7)	(1.96)	(0.28)	(9.2)
计	14	51.65		100.0

带○号的为并入误差项。

J 对于收率来说,C 和 H 对凝固点来说是有疑问的因素,即使接近 5%,也要计算平均值,如果成本相同,应该根据水平选优。在几个特性值的情况下,估计值的一览见表 46.3-30。带○号表示按  $\alpha=1\%$  判断为好;○号表示按  $\alpha=5\%$  判断时为好;△号表示无显著效应。但是按“似乎是好的”考虑。对于凝固点 B、D 的主效应尽管不大,但是 B×D 的效应非常大。调查 E、F 是否存在交互作用,E×F 在第五列,这时第五列为 0,是把欠测值代入的结果。误差自由度太小时,称为残差平方和最小化。欠测值的代入有时会发生大的错误,代

入平均值用逐次估计是很好的方法。

表 46.3-30 分析结构汇总表

因素	水平		收率提高	凝固点
A 原料油	○1	58 轻	+5.0○	-16.0○
	○2	65 重	+1.0	15.2
B 苯使用量	○1	22%	-	-
	2	24%	-	-
AB	11		+6.3○	-
	12		+3.7	-
	22		-1.9○	-
	21		+0.0	-
C 调配罐温度	○1	30°C	-	-15.8△
	2	35°C	-	-15.3
D 原料油,溶剂	○1	1:3.8	+1.2	-16.0○
	2	1:4.6	+4.8○	-15.2
BD	11		+0.05	-16.7○
	12		+5.8○	-14.2
	21		+1.9	-15.3
	22		+3.7	-16.2
E 冷却温度	1	-15°C	-	-14.4
	○2	-17°C	-	-16.7○
F 冷却速度	1	3 台	-	-16.4○
	○2	2 台	-	-14.8
G 喷嘴位置	○1	现行	+2.2	-
	2	缩小 1/10	+3.8○	-
H 离心分离机台数	○1	9 台	+4.4○	-15.9△
	2	8 台	+1.6	-15.3
I 补充溶剂温度	○1	-20°C	-	-
	2	-17°C	-	-
J 补充溶剂流量	1	3.3kL/h	+2.3	-15.9○
	○2	2.8kL/h	+3.7△	-15.3
置界限	主效应		+1.6	±0.63
	交互作用		±2.2	±0.86

注:表中,水平是用配列表指定的水平,一部分替换最初水平。水平栏的○号表示当时的现行水平。

### 3.5 结论

由表 46.3-30 可知,D、B×D 以及 J 外的因素只对一个特性有好作用,而 H 的第一水平对两个特性都有好的作用。特别是 E、F 对凝固点有很好的作用。所以对 E 取现行水平 -17°C,对 F 取新水平(使用 3 台冷却器),然后可以选取有利于收率的方案。

B:苯的使用量,取现行的 22%。

C:选现行的 30°C。

D:不管 B 如何应该选用 1:4.6。

E:应该绝对地选用 E<sub>2</sub> 的 -17°C。

F:应该使用 3 台。

G:缩小 1/10。

H:使用 9 台。

I:补充溶剂任选,原则上选现行的。

J:哪一种水平不能限定,一般选 3.3kL/h。

由于对应于 A<sub>1</sub>、A<sub>2</sub>,保留了 B<sub>1</sub>、B<sub>2</sub>,所以要估计 B<sub>1</sub>C<sub>1</sub>D<sub>2</sub>E<sub>2</sub>F<sub>1</sub>G<sub>2</sub>H<sub>1</sub>I<sub>1</sub>J<sub>1</sub> 和 B<sub>2</sub>C<sub>1</sub>D<sub>2</sub>E<sub>2</sub>F<sub>1</sub>G<sub>2</sub>H<sub>1</sub>I<sub>1</sub>J<sub>1</sub> 的平均工序。

对于  $A_1B_1C_1D_2E_2F_1G_2H_1I_1J_1$ , 忽略非重要因素后, 根据表 46.3-17 可以算出: 收率的工序平均为

$$\begin{aligned} \bar{\mu} &= (A_1B_1) + (B_1D_2) - 2(B_1) + (G_2) + (H_1) - (J_1) \\ &\quad - 2(\text{平均}) \\ &= 6.3 - 5.8 - 2\left(\frac{25.1}{8}\right) + 3.8 + 4.4 + 2.3 - 2 \times 3.0 \\ &= -10.4 \pm 2.2 \end{aligned}$$

凝固点的工序平均

$$\begin{aligned} \bar{\mu} &= (A_1) + (C_1) + (B_1D_2) - (B_1) + (E_2) + (F_1) \\ &\quad + (H_1) - (J_1) - 5(\text{平均}) \\ &= -16.0 - 15.8 - 14.2 + 15.4 - 16.7 - 16.4 \\ &\quad - 15.9 + 15.6 - 5 \times (-9.3) = -17.5 + 1.0 \end{aligned}$$

置信界限公式:

$$\pm \sqrt{F(0.05) \times V_r \times \frac{1}{n_r}}$$

对于收率

$$\begin{aligned} n_r &= \frac{16}{\bar{\mu} \text{各因素的自由度之和}} \\ &= \frac{16}{A \times B, B \times D \text{ 和 } A, D, G, H, J, m \text{ 的自由度和}} \\ &\quad - \frac{16}{8} \\ F(0.05) &= 5.59 \\ V_r &= 1.71 \end{aligned}$$

得  $\pm \sqrt{5.59 \times 1.71 \times \frac{8}{16}} = \pm 2.2$

对于凝固点

$$\begin{aligned} n_r &= \frac{16}{m, A, C, D, E, F, H, J \text{ 和 } B \times D \text{ 的自由度之和}} \\ &= \frac{16}{9} \\ F(0.05) &= 5.59 \\ V_r &= 0.28 \end{aligned}$$

得

$$\pm \sqrt{5.59 \times 0.28 \times \frac{9}{16}} = \pm 1.0$$

同样估计  $A_1B_2 \dots, A_2B_1 \dots, A_2B_2 \dots$  的工序平均, 得到表 46.3-31。

表 46.3-31

运输条件	收率	比现在提高的量	凝固点	与现行比较
$A_1B_1C_1D_2E_2F_1G_2H_1I_1$	10.4 -2.2	+5.7%	-17.5 ±1.0	-0.1°C
$A_1B_2C_1D_2E_2F_1G_2H_1I_1$	6.0 ±2.2	-1.3%	-19.3 ±1.0	-1.7°C
$A_2B_1C_1D_2E_2F_1G_2H_1I_1$	4.2 ±2.2	+6.1%	-16.5 ±1.0	-0.3°C
$A_2B_2C_1D_2E_2F_1G_2H_1I_1$	4.2 ±2.2	+6.1%	-18.5 ±1.0	-1.7°C

与现行比较是与  $B_1C_1D_1E_2F_2G_2H_1I_1J_1$  的工序平均比较。现行条件的预计收率值为表 46.3-32;

表 46.3-32

	收率	凝固点
$A_1$	(4.7±2.3)%	(-17.6±1.0)°C
$A_2$	(-1.9±2.3)%	(-16.8±1.0)°C

实际上用  $A_2, B_1, C_1, D_2, E_2, F_1, I_1, J_1$  进行验证运转的结果为

收率 4.1%  
凝固点 -15.4°C

收率恰好进入预定的  $4.2 \pm 2.2$  内, 凝固点之所以没有同预计值  $-16.5 \pm 1.0$  °C, 是因为预计值的置信界限是多次验证运转的平均估计, 而实际只运转了几次。如果是  $r$  次的平均时, 以 95% 的置信度, 应该进入

$$-16.5 \pm \sqrt{F \times V_r \times \left(\frac{1}{n_r} + \frac{1}{r}\right)} \text{ 内的,}$$

令  $r=3, n_r=16/9$  得

$$-16.5 \pm \sqrt{5.59 \times 0.28 \times \left(\frac{9}{16} + \frac{1}{3}\right)} = -16.5 \pm 1.2$$

此时确实进入置信界限。

## 4 羊毛纺纱中的清洗过程和梳理过程实验

### 4.1 目的

原毛分选后, 要进行清洗, 用梳毛机进行梳毛。为决定洗毛与梳毛工序各因素的最宜作业标准, 要进行调查这些工序因素对纤维的损伤、损失、纺织性能等影响的实验。

### 4.2 因素与水平

洗毛工序的因素水平:

- X: 残脂率;
- Y: 水分;
- Z: 给油量。

梳毛工序的因素水平

- A: 梳毛供给量;
- B: 送料滚筒隔距;
- C: 反毛机工作辊隔距;
- D: 工作辊隔距;
- E: 第一提毛辊隔距;
- F: 第二提毛辊隔距;
- G: 反毛机转速;
- H: 除尘底格隔距;
- I: 工作辊转速;
- J: 第一提毛辊转速;

K:第二提毛辊转速;  
L:落纱机转速;  
M:传送装置转速。  
以上因素全为二水平。

4.3 数据与配列

把洗毛的三个因素 X、Y、Z 对应于正交表  $L_4(2^3)$  的三个列见表 46.3-33,把梳毛工序的 13 个因素 A~M 配列于正交表  $L_{16}(2^5)$ (见表 46.3-34),两方组合,以随机顺序进行  $4 \times 16 = 64$  次实验。

表 46.3-33

	X	Y	Z
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

因素 A~M 间的交互作用,在决定水平时考虑了,配列时未考虑。为了能全面的分析,洗毛工序因素与梳毛工序因素间的交互作用进行了全面的组合实验,即进行了的  $L_4 \otimes L_{15}$  直积。对难以改变水平的因素进行了随机化,进行测量的特性值近 10 种,也做了综合判定,本例只给出了毛条的毛结数。

表 46.3-34 配列与毛结数

	A	B	C	D	E	F	G	H	I	J	K	L	M	X <sub>1</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>2</sub>
	1	2	14	4	13	11	7	15	8	6	12	9	10	Y <sub>1</sub>	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>2</sub>
1	1	1	1	1	1	1	1	1	1	1	1	1	1	18	14	18	16
2	1	1	2	1	2	2	1	2	2	1	2	2	2	19	16	18	21
3	1	1	2	2	2	1	2	2	1	2	2	1	1	33	24	26	32
4	1	1	1	2	1	2	2	1	2	2	1	2	2	22	19	24	21
5	1	2	2	1	1	2	2	2	1	2	1	1	2	16	13	17	17
6	1	2	1	1	2	1	2	2	1	2	2	2	1	17	14	17	19
7	1	2	1	2	2	2	1	1	1	1	2	1	2	28	28	25	26
8	1	2	2	2	1	1	1	2	2	1	1	2	1	21	17	21	20
9	2	1	1	1	2	2	2	2	1	1	1	2	1	17	14	16	17
10	2	1	2	1	1	1	2	1	2	1	2	1	2	21	18	21	18
11	2	1	2	2	1	2	1	1	1	2	2	2	1	34	28	25	35
12	2	1	1	2	2	1	1	2	2	2	1	1	2	21	18	19	21
13	2	2	2	1	2	1	1	1	1	2	1	2	2	16	13	18	20
14	2	2	1	1	1	2	1	2	2	2	2	1	1	25	21	22	22
15	2	2	1	2	1	1	2	2	1	1	2	2	2	20	24	29	24
16	2	2	2	2	2	2	2	1	2	1	1	1	1	27	19	26	26

在缺陷数多的部位取单位面积,可以在形成毛条前,从中心与左、右两部分,每小时取样 2~3 点共 6~9 点,求出数据,尽可能分析各因素对毛结数有什么样的影响,以及如何随时间变化。

4.4.2 分散指数法

与二项分布中  $x^2$  检验法完全相同。当单位面积取得过大时,此法的分析精度高。当因素效应有特大效应的时候,或者显著因素很多时,一般方差分析法可能得到高的分析精度。一般对  $x^2$  检验法来说,相当于误差方差的成分中,因素效应全包括了。所以一般不用此法。

4.4.3 变数变换法

此法计算比较麻烦,精度不高的情况较多,通常取数据的平方根或者对数值,作计量值的方差分析。

毛结数的合计,可按各因素的每一水平求出,但在本例情况下,需要求 X、Y、Z 与 A~M 之间的交互作用,所以作辅助表(见表 46.3-35)。

表 46.3-35 辅助表

	(1) X <sub>1</sub> Y <sub>1</sub> Z <sub>1</sub>	(2) X <sub>1</sub> Y <sub>2</sub> Z <sub>2</sub>	(3) X <sub>2</sub> Y <sub>1</sub> Z <sub>2</sub>	(4) X <sub>2</sub> Y <sub>2</sub> Z <sub>1</sub>	计
A	174	145	166	172	657
A	187	158	176	183	704
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
B <sub>1</sub>	185	151	167	181	684
B <sub>2</sub>	176	152	175	174	677
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
C <sub>1</sub>	174	155	170	166	665
C <sub>2</sub>	187	148	172	189	696
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
D <sub>1</sub>	149	123	147	150	569
D <sub>2</sub>	212	180	195	205	792
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
E <sub>1</sub>	183	157	177	173	690
E <sub>2</sub>	178	146	165	182	671
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
F <sub>1</sub>	173	145	169	170	657
F <sub>2</sub>	188	158	173	185	704
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
G <sub>1</sub>	182	155	166	181	684
G <sub>2</sub>	179	148	176	174	677
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
H <sub>1</sub>	183	153	174	181	691
H <sub>2</sub>	178	150	168	174	670
计	361	303	342	355	1361

4.4 数据分析

缺陷数(毛结数、断条、故障数)的分析。

4.4.1 一般的方差分析

为了有效地运用一般方差分析方法,必须合理选择计数缺陷数的单位面积,按各实验单位分割检验面积。各细分面积所摊的缺陷数的平均值,大体上应该有一个至数个。本例的数据平均有 20 个,应该把检验面积数等分,记录细分面积上的缺陷数,不作合计。如果



(续)

	(1)	(2)	(3)	(4)	计
	$X_1Y_1Z_1$	$X_1Y_2Z_2$	$X_2Y_1Z_2$	$X_2Y_2Z_1$	
$I_1$	188	161	174	187	710
$I_2$	173	142	168	168	651
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
$J_1$	177	153	174	168	672
$J_2$	184	150	168	187	689
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
$K_1$	158	127	159	158	602
$K_2$	203	176	183	197	759
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
$L_1$	189	155	174	178	696
$L_2$	172	148	168	177	665
计	361	303	342	355	1361
	(1)	(2)	(3)	(4)	计
$M_1$	192	151	171	187	701
$M_2$	169	152	171	168	660
计	361	303	342	355	1361

用分散指数法,  $x^2$  的值可用下式计算:

当  $A$  有  $a$  个水平时,

$$x_A^2 = \frac{a}{A_1 + \dots + A_a} \left[ \left( A - \frac{A_1 + \dots + A_a}{a} \right)^2 + \dots + \left( A_a - \frac{A_1 + \dots + A_a}{a} \right)^2 \right]$$

当  $A$  为二水平、因素的自由度为 1 时, 可以求得

$$x_A^2 = \frac{(A_1 - A_2)^2}{A_1 + A_2} = \frac{(657 - 704)^2}{1361} = 1.6$$

效应显著的因素非常多, 如果用  $x^2$  检验, 精度较低, 所以采用方差分析法。

$$S_A = \frac{(A_1 - A_2)^2}{64} = \frac{47^2}{64} = 34.5$$

$$S_X = \frac{(361 + 303 - 342 - 355)^2}{64} = 17.0$$

$$S_Y = \frac{(361 + 342 - 303 - 355)^2}{64} = 31.6$$

$$S_Z = \frac{(361 + 355 - 342 - 303)^2}{64} = 78.8$$

对于交互作用, 先分别就(1)、(2)、(3)、(4)求  $A_1$ 、 $A_2$  之差, 然后进行计算较快。

$$A_1 - A_2 \quad (1) \quad (2) \quad (3) \quad (4)$$

$$-13 \quad -13 \quad -10 \quad -11$$

$$S_{A \times X} = \frac{(-13 - 13 + 10 + 11)^2}{64} = 0.4$$

$$S_{A \times Y} = \frac{(-13 - 10 + 13 + 11)^2}{64} = 0.0$$

$$S_{A \times Z} = \frac{(-13 - 11 + 10 + 13)^2}{64} = 0.0$$

其他因素的计算方法相同。结果如表 46.3-36 所示。

表 46.3-36 方差分析表

因素	$f$	$S$	因素	$f$	$S$	因素	$f$	$S$
$X$	1	17.0	$B \times X$	1	1.3	$H \times Y$	1	0.0
$Y$	1	31.6	$B \times Y$	1	0.4	$H \times Z$	1	0.1
$Z$	1	78.8	$B \times Z$	1	9.8	$I \times X$	1	1.3
$A$	1	34.5	$C \times X$	1	5.6	$I \times Y$	1	4.5
$B$	1	0.8	$C \times Y$	1	0.0	$I \times Z$	1	1.3
$C$	1	15.0	$C \times Z$	1	26.3	$J \times X$	1	1.3
$D$	1	777.0	$D \times X$	1	4.5	$J \times Y$	1	3.5
$E$	1	5.6	$D \times Y$	1	0.0	$J \times Z$	1	19.1
$F$	1	34.5	$D \times Z$	1	2.6	$K \times X$	1	15.0
$G$	1	0.8	$E \times X$	1	2.6	$K \times Y$	1	5.6
$H$	1	6.9	$E \times Y$	1	3.5	$K \times Z$	1	1.9
$I$	1	54.4	$E \times Z$	1	11.4	$L \times X$	1	4.5
$J$	1	4.5	$F \times X$	1	1.8	$L \times Y$	1	3.5
$K$	1	385.1	$F \times Y$	1	1.3	$L \times Z$	1	0.4
$L$	1	15.0	$F \times Z$	1	2.8	$M \times X$	1	0.1
$M$	1	26.3	$G \times X$	1	2.6	$M \times Y$	1	0.4
$A \times X$	1	0.4	$G \times Y$	1	6.9	$M \times Z$	1	28.7
$A \times Y$	1	0.0	$G \times Z$	1	2.6	$e$	8	43.2
$A \times Z$	1	0.0	$H \times X$	1	0.4	计	63	1708.5

效应小的因素合并为(e)

(48)

(150.44)

(3.18)

### 4.5 结论

本例中仅就毛结数求最宜条件。从交互作用的显著性看, 好的组合为:

$C, Z$  的组合为  $C_2Z_2$

$J, Z$  的组合为  $J_2Z_2$

$M, Z$  的组合为  $M_1Z_2$

$K, Z$  的组合为  $K_1Z_1$

在这种情况下,  $Z_2$  普遍好, 所以容易决定。仅主效

应显著的  $Y, A, D, F, J, L$ , 根据辅助表容易确定其好的水平为  $Y_2, A_1, D_1, F_1, J_2, L_2$ 。因而最宜条件为  $Y_2, A_1, D_1, F_1, J_2, L_2$ 。

其工序平均为:

$$\begin{aligned} \hat{\mu} &= \frac{C_2Z_2}{16} + \frac{J_2Z_2}{16} + \frac{M_1Z_2}{16} + \frac{K_1X_1}{16} - 2 \frac{Z_2}{32} - \frac{J_2}{32} \\ &+ \frac{Y_2}{32} + \frac{A_1}{32} + \frac{D_1}{32} + \frac{F_1}{32} + \frac{I_2}{32} + \frac{L_2}{32} - 6 \frac{T}{64} \\ &= \frac{320}{16} + \frac{318}{16} + \frac{322}{16} + \frac{285}{16} - 2 \times \frac{645}{32} - \frac{689}{32} + \frac{658}{32} \end{aligned}$$

$$= \frac{657}{32} + \frac{569}{32} + \frac{657}{32} + \frac{651}{32} + \frac{665}{32} - 6 \times \frac{1361}{64}$$

$$= \frac{285}{32} \approx 8.9$$

有效反复数  $n_e$  为

$$n_e = \frac{\text{实验总数}}{\text{不能忽视的因素的自由度之和}} = \frac{64}{16} = 4$$

置信界限为:

$$x = \hat{\mu} \pm \sqrt{F \times V_e \times \left( \frac{1}{n_e} + 1 \right)}$$

$$= 8.9 \pm \sqrt{4.04 \times 3.18 \times \left( \frac{1}{4} + 1 \right)}$$

$$= 8.9 \pm 4.0$$

对应于现行作业标准  $X_2Y_2Z_2A_1C_2D_1E_2I_1J_1K_2L_1M_1$  的工序平均为

$$\hat{\mu} = \frac{C_2X_2}{16} + \frac{J_1Z_2}{16} + \frac{M_1Z_2}{16} + \frac{K_2X_2}{16} - 2 \frac{X_2}{32} - \frac{J_1}{32} + \frac{Y_2}{32}$$

$$+ \frac{A_1}{32} + \frac{D_1}{32} + \frac{F_2}{32} + \frac{I_2}{32} + \frac{J_1}{32} - 6 \frac{T}{64}$$

$$= \frac{320}{16} + \frac{327}{16} + \frac{322}{16} + \frac{380}{16} - 2 \frac{645}{32}$$

$$- \frac{672}{32} + \frac{658}{32} + \frac{657}{32} + \frac{569}{32} + \frac{704}{32}$$

$$+ \frac{710}{32} + \frac{696}{32} - 6 \frac{1361}{64}$$

$$= \frac{647}{32} = 20.218$$

## 5 聚酯纤维混纺布料的生产条件

### 5.1 因素与水平

为决定涤纶与羊毛混纺的制造条件,选取了表 46.3-37 的因素与水平。 $C_3$  水平只在  $A_1, A_2$  水平下可以制造,在  $A_3, A_4$  水平下不能制造。 $C_1, C_2$  水平在  $A_1, A_2, A_3, A_4$  任一水平下均能制造。 $I, J$  是热定型的方法,是仅对  $H_2$  水平有意义的因素。还有密度的大中小根据支数与股数决定,如表 46.3-37 下部所示。

表 46.3-37 因素与水平

因素	记号	第一水平	第二水平	第三水平	第四水平
质量改进的涤纶种类	A	$A_1$	$A_2$	$A_3$	$A_4$
混纺率(涤纶、羊毛)	B	55 : 45	45 : 55		
支数	C	2/60 普通	2/52 特殊	2/72	
终拈数	D	(67s/55z, 71s/59z, 77s/65z)	(90s/55z, 107s/59z, 117s/66z)		
染色方法	E	染纤维条	染线		
织物组织	F	绫	平斜纹	鸟眼纹	
织物密度	G	大	中	小	
热定形的有无	H	无	有		
热定型的方法	I	圆网	拉幅机	仅对 $H_2$ 有意义	
热定型工序的位置	J	洗绒前	洗绒后	仅对 $H_2$ 有意义	
取绒毛方法	K	剪毛	煤气烧		
精整定型方法	L	转塔	罐蒸绒	蒸绒	
洗绒方法	M	HH	绳状整理	扩幅	

因素的相互关系(密度)

普通拈度	大(%)	中(%)	小(%)	特别拈度	大(%)	中(%)	小(%)
2/52	108.0	100.0	97.5	2/52	101.0	98.0	95.0
2/60	103.0	100.0	97.5	2/60	100.5	97.5	95.0
2/72	102.5	100.0	97.5	2/72	100.0	97.5	95.0

注:东纺、大同毛织的共同研究(1965)。

### 5.2 配列与数据

作出点线图(见图 46.3-4),配列于  $L_{32}$  正交表(见表 46.3-38)。进行了手感、针通、精整幅宽、幅 SK(%)、

精整长度 SK(%)平方米重量、经线收缩、纬线收缩、皱纹回复率、弹性率、磨耗强度及其它等数据的测量。本节只考虑手感的数据由  $R_1, R_2, R_3, \dots, R$  判定的。数据如表 46.3-38 所示。

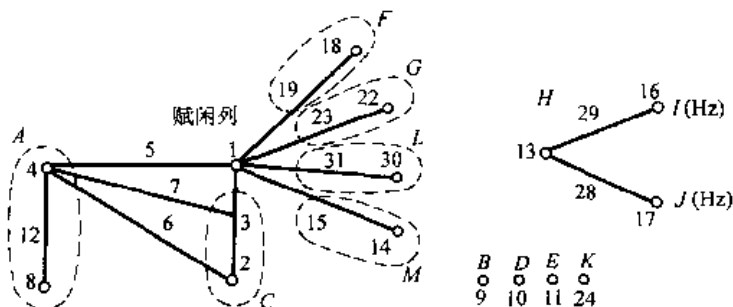


图 46.3-4 点线图



$$S_{(C_2, C_1)} = \frac{(3-7)^2 \times 4.90 + (17-22)^2 \times 5.27}{56}$$

$$= 3.75 (f=2)$$

$$S_D = 25.00 (f=2)$$

$$S_E = 6.08 (f=2)$$

$$S_{(F_1, F_2)} = \frac{(25-19)^2 \times 4.90 + (53-47)^2 \times 5.27}{112}$$

$$= 3.27 (f=2)$$

$$S_{(F_1, F_3)} = \frac{(15-5)^2 \times 4.90 + (38-29)^2 \times 5.27}{112}$$

$$= 8.19 (f=2)$$

$$S_{(G_1, G_2)} = 0.89 (f=2)$$

$$S_{(G_2, G_3)} = 2.00 (f=2)$$

$$S_H = 3.18 (f=2)$$

$$S_{(L_1, L_2)} = 9.08 (f=2)$$

$$S_{(L_2, L_3)} = 3.98 (f=2)$$

$$S_{(M_1, M_2)} = 3.56 (f=2)$$

$$S_{(M_2, M_3)} = 1.18 (f=2)$$

$$S_K = 7.14 (f=2)$$

根据  $H_2$  的数据得

$$S_I = \frac{(20-18)^2 \times 4.9 + (41-42)^2 \times 5.27}{224}$$

$$= 0.11 (f=2)$$

$$S_J = 0.11 (f=2)$$

$$S_M = \frac{(44-20)^2 \times 4.9 + (100-67)^2 \times 5.27}{224}$$

$$= 38.41 (f=2)$$

$$S_{T_1} = \frac{(5+1^2+0^2+\dots+2^2) \times 4.9 +$$

$$\frac{(7^2+5^2+5^2+\dots+6^2) \times 5.27}{7} - CF$$

$$= 143.24 (f=2)$$

$$S_{T_1} = S_T - (S_A + S_B + \dots + S_M) = 143.24 - 125.87$$

$$= 17.37 (f=2)$$

$$S_R = \frac{(17^2+6^2+\dots+2^2) \times 4.9 +$$

$$\frac{(32^2+24^2+\dots+20^2) \times 5.27}{7} - CF$$

$$= 38.44 (f=12)$$

$$S_{T_2} = S_T - S_{T_1} - S_R = 148 - 143.24 - 38.44$$

$$= 266.32 (f=372)$$

由此得到方差分析表 46.3-40。

表 46.3-40

因素	f	S	V	F <sub>0</sub>
赋闲	2	38.41	19.20	27.0
A	6	8.29	1.38	1.93
B	2	0.68	0.34	—
C <sub>1</sub> C <sub>2</sub>	2	0.97	0.48	—
C <sub>2</sub> C <sub>3</sub> (A <sub>1</sub> A <sub>2</sub> )	2	3.75	1.88	2.64
D	2	25.00	12.50	17.56
E	2	6.08	3.04	4.27
F <sub>1</sub> F <sub>2</sub>	2	3.27	1.64	2.30
F <sub>1</sub> F <sub>3</sub>	2	8.19	4.10	5.76
G <sub>1</sub> G <sub>2</sub>	2	0.89	0.44	—
G <sub>2</sub> G <sub>3</sub>	2	2.00	1.00	—
H	2	3.18	1.59	2.23
L <sub>1</sub> L <sub>2</sub>	2	9.08	4.54	6.33
L <sub>2</sub> L <sub>3</sub>	2	3.98	1.99	2.79
M <sub>1</sub> M <sub>2</sub>	2	3.56	1.78	2.50
M <sub>2</sub> M <sub>3</sub>	2	1.18	0.59	—
K	2	7.14	3.57	5.01
I	2	0.11	0.06	—
J	2	0.11	0.06	—
e <sub>1</sub>	20	17.37	0.87	—
R	12	38.44	3.20	4.49
e <sub>2</sub>	372	266.32	0.716	—
计	446	448.00		
合并(e)	(406)	(289.05)	(0.712)	
(e)	(422)	(315.66)	(0.748)	

### 5.3.2 估计

由方差分析表得出,效应显著得因素有7个。赋闲列中由于有各种因素混杂,实际上不称其为因素。因而包括一般平均的估计,如表 46.3-41 所示。由表 46.3-40 数据画成曲线图 46.3-5。

表 46.3-41

	I	II	III	上	中	下	计	(% )			计
								上	中	下	
D <sub>1</sub>	47	91	112	47	41	21	112	42.0	39.2	18.8	100.0
D <sub>2</sub>	17	76	112	17	59	36	112	15.1	52.8	32.1	100.0
E <sub>1</sub>	35	76	112	35	41	36	112	31.2	36.7	32.1	100.0
E <sub>2</sub>	29	91	112	29	62	21	112	25.9	55.3	18.8	100.0
F <sub>1</sub>	15	38	56	15	23	18	56	26.8	41.1	32.1	100.0
F <sub>2</sub>	5	29	56	5	24	27	56	8.9	42.9	48.2	100.0
L <sub>1</sub>	27	55	56	27	28	1	56	48.2	50.0	1.8	100.0
L <sub>2</sub>	17	45	56	21	28	11	56	30.4	50.0	19.6	100.0
K <sub>1</sub>	39	89	112	39	50	23	112	34.8	44.7	20.5	100.0
K <sub>2</sub>	25	78	112	25	53	34	112	22.3	47.3	30.4	100.0
计	64	167	224	64	103	57	224	28.6	46.0	25.4	100.0

置信界限近似计算:

$$\pm \sqrt{F \times V \times \mu(1-\mu) \times \frac{1}{n}}$$

对应于 D 的上组和(上+中)组的置信界限为

$$= \sqrt{3.86 \times 0.748 \times 0.42 \times 0.58 \times \frac{1}{112}} = \pm 0.079$$

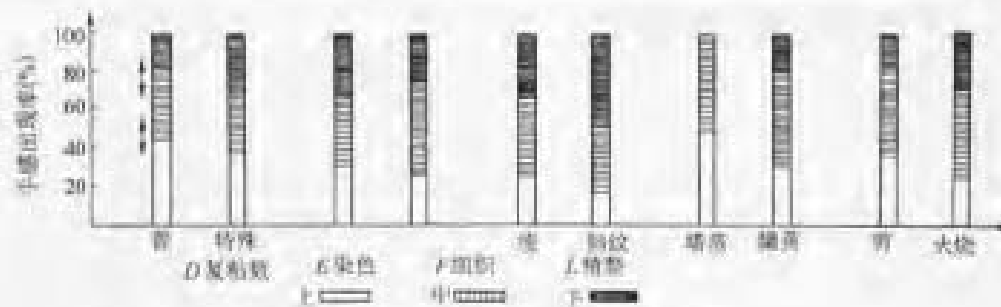


图 46.3-5 效应因素曲线图

$$\pm \sqrt{3.86 \times 0.748 \times 0.812 \times 0.188 \times \frac{1}{112}}$$

$$= \pm 0.063$$

### 5.3.3 最宜条件与工序平均

根据上节数据求有关手感的最宜条件,得到  $D_1, E_1, F_1, L_1, K_1$  或  $D_1, E_1, F_1, L_1, K_1$ 。工序平均的计算用欧米加法,先导出算术可加性的公式。

$$\bar{\mu} = \bar{T} + (\bar{D}_1 - \bar{T}) + (\bar{E}_1 - \bar{T}) + (\bar{K}_1 - \bar{T}) +$$

$$(\bar{F}_1 - \bar{T}) + (\bar{L}_1 - \bar{T}) - \frac{1}{2}$$

(赋闲 1 的列  $\bar{L}_1$  - 赋闲 2 的平均)

$$= \bar{D}_1 + \bar{E}_1 + \bar{K}_1 + \bar{F}_1 + \bar{L}_1 - 4\bar{T} - \frac{1}{2}$$

(赋闲 1 的列  $\bar{L}_1$  - 赋闲 2 的平均)

式中  $\bar{F}_1$  是通过赋闲 1、2 两方的平均,即修正后的平均。因此选  $\bar{E}_1$  时

$$\bar{\mu}_1 \text{ 的分贝值} = -1.40 - 3.43 - 2.73 - 2.55 -$$

$$0.31 - 4(-3.97) - \frac{1}{2}(-3.6$$

$$+ 6.61) = 6.90(\text{dB})$$

$$\bar{\mu}_2 \text{ 的分贝值} = 6.35 + 3.25 + 3.89 + 6.35 + 17.37$$

$$- 4 \times 4.68 - \frac{1}{2}(0.00 - 1.72)$$

$$= 21.35(\text{dB})$$

在选  $E_1$  时,把  $(\bar{E}_1 - \bar{E}_2)$  的分贝值,加入上式值

$$\bar{\mu}_1 \text{ 的分贝值} = 5.83\text{dB}$$

$$\bar{\mu}_2 \text{ 的分贝值} = 24.45\text{dB}$$

1、1 组的分贝值如表 46.3-42 所示。

表 46.3-42

因素	1 组	分贝值	1 组	分贝值
$D_1$	42.0	-1.40	81.2	5.30
$E_1$	21.2	-3.43	67.9	3.25
$K_1$	25.9	-4.56	81.2	6.35
$F_1$	34.8	-2.73	79.3	5.89
$L_1$	35.7	-2.55	81.2	6.35

(续)

因素	1 组	分贝值	1 组	分贝值
$L_1$	48.2	-0.31	98.2	17.57
赋闲 $L_1$	30.4	-3.60	50.0	6.00
赋闲 2	17.9	-6.61	58.8	1.72
$T$	28.6	-3.97	74.6	0.68

得可供选择得最宜条件(如表 46.3-43 所示)。

表 46.3-43

条件	1	1	1	上	中	下	计
$D_1, E_1, K_1, F_1, L_1$	83.2	99.4	100.0	83.2	16.2	0.0	100.0
$D_1, E_1, K_1, F_1, L_1$	79.3	98.6	100.0	79.3	29.2	0.4	100.0

最宜条件  $D_1, E_1, K_1, F_1, L_1$ , 生产 100 匹时的手感预测值为

$$\text{上组} = 83.2 \pm$$

$$\sqrt{3.85 \times 0.748 \times 0.832 \times 0.188 \left( \frac{19}{448} + \frac{1}{100} \right)}$$

$$\times 100 = 83.2 \pm 14.5(\%)$$

(上+中)组

$$= 99.4 \pm$$

$$\sqrt{3.85 \times 0.748 \times 0.994 \times 0.006 \left( \frac{19}{448} + \frac{1}{100} \right)}$$

$$\times 100 = 99.4 \pm 3.0(\%)$$

其中  $n_1$

$$1 = \frac{\text{考虑了工序平均的因素实验自由度之和}}{\text{实验的总自由度}}$$

$$n = D \quad E \quad K \quad F \quad L_1$$

$$= \frac{2+2+2+2+9}{448} = \frac{19}{448}$$

## 6. 密封工序和灯泡外观品质的实验

### 6.1 目的

在灯泡生产厂的密封工序是把心棒插入玻璃管中,进行密封,密封后灯泡发生大量形状不良,且造成增加筛选费用和较大的不良率,为此,以密封工序为中心的实验,调查是否可以改进形状不良的问题,同时为了提高密封工序的生产率,调查提高密封速度对形状不良率的影响。

6.2 因素与水平

玻璃管的壁厚不均对形状不良有影响,但是由于外采购而且缺乏相关检测设施,未作为因素。提出的因素为以下:

A:心棒直径 按心棒的平均值特制二分级的通止塞规,将心棒直径分为两组。

A<sub>1</sub>:直径大的

A<sub>2</sub>:直径小的

B:密封机速度

B<sub>1</sub>:4.80r/min

B<sub>2</sub>:540r/min

C } 第一阶段的瓦斯量因素 C 与空气量因素 D

E } 第二阶段的瓦斯量因素 E 与空气量因素 F

G } 第三阶段的瓦斯量因素 G 与空气量因素 H

I } 第四阶段的瓦斯量因素 I 与空气量因素 J

玻璃式自动通过预热、加热、密封、切断等阶段,装入密封机,从这个阶段中,只提出了似乎与形状不良有关的包括密封工序的四个阶段。各个阶段的幅度控制是通过调节瓦斯量与空气量的阀门进行的。每一阶段是独立控制的,所以把调节四个阶段的瓦斯量与空气量的阀门位置取为因素。各因素的水平决定如下:

C<sub>1</sub>:瓦斯量的标准位置 { D<sub>1</sub>:标准位置  
D<sub>2</sub>:从标准位置减少一个量度

C<sub>2</sub>:增加一个量度的瓦斯量 { D<sub>1</sub>:从标准位置增加一个量度  
D<sub>2</sub>:标准位置

(E,F)、(G,H)、(I,J)也同样。

上述因素可以多次随机改变水平。重要的因素是第三、四阶段,特别是 G×H 和 I×J,希望尽可能不要同其他因素的主效应混杂,并将灯泡分成三种:完全良品,合格的形状不良品与不良品。

6.3 配列与实验结果

为使 G×H 和 I×J 出现,采用 L<sub>16</sub>(2<sup>15</sup>)正交表。用 B 表示不良品,用 T 表示合格的形状不良品,用 G 表示完全良品。每次实验制造了 10 个灯泡,10 个灯泡中的 G、I、B 的数量如表 46.3-44 所示。

表 46.3-44

	A	B	C	D	E	F	G	H	I	J	G	T	B
1	1	1	1	1	1	1	1	1	1	1	5	4	1
2	1	1	2	2	2	2	1	1	1	2	6	1	0
3	2	2	1	1	2	2	1	1	2	1	4	4	2

(续)

	A	B	C	D	E	F	G	H	I	J	G	T	B
4	2	2	2	2	1	1	1	1	2	2	3	5	2
5	2	2	2	2	1	2	1	2	1	1	2	3	5
6	2	2	1	1	2	1	1	2	1	2	6	3	1
7	1	1	2	2	2	1	1	2	2	1	3	0	7
8	1	1	1	1	1	2	1	2	2	2	2	2	6
9	1	2	1	2	2	1	2	1	1	1	4	4	2
10	1	2	2	1	1	2	2	1	1	2	5	5	0
11	2	1	1	2	1	2	2	1	2	1	4	4	2
12	2	1	2	1	2	1	2	1	2	2	3	3	4
13	2	1	2	1	2	2	2	2	1	1	2	4	4
14	2	1	1	2	1	1	2	2	1	2	8	2	0
15	1	2	2	1	1	1	2	2	2	1	2	3	5
16	1	2	1	2	2	2	2	2	2	2	2	1	7

L<sub>16</sub>列号 6 7 10 11 13 14 1 2 4 8  
I×J→12列, G×H→3列, e→5,9,15列

6.3.1 累计法分析,方差分析

设 I=C, II=G+T, III=G+T+B,作累积频数辅助表(见表 46.3-45)。

首先求 I、II 组的权数 W<sub>1</sub>、W<sub>2</sub>。

$$W_1 = \frac{160^2}{61(160-61)} = 4.2391$$

$$W_2 = \frac{160^2}{112(160-112)} = 4.7619$$

表 46.3-45 辅助表

	I	II	III		I	II	III
A <sub>1</sub>	29	52	80	B <sub>1</sub>	33	56	80
A <sub>2</sub>	32	60	80	B <sub>2</sub>	28	56	80
计	61	112	160	计	61	112	160
	I	II	III		I	II	III
C <sub>1</sub>	35	59	80	D <sub>1</sub>	29	57	80
C <sub>2</sub>	26	53	80	D <sub>2</sub>	32	55	80
计	61	112	160	计	61	112	160
	I	II	III		I	II	III
E <sub>1</sub>	31	59	80	F <sub>1</sub>	34	58	80
E <sub>2</sub>	30	53	80	F <sub>2</sub>	27	54	80
计	61	112	160	计	61	112	160
	I	II	III		I	II	III
G <sub>1</sub>	31	56	80	H <sub>1</sub>	34	67	80
G <sub>2</sub>	30	56	80	H <sub>2</sub>	27	45	80
计	61	112	160	计	61	112	160
	I	II	III		I	II	III
(G×H) <sub>1</sub>	32	59	80	I <sub>1</sub>	38	67	80
(G×H) <sub>2</sub>	29	53	80	I <sub>2</sub>	23	45	80
计	61	112	160	计	61	112	160
	I	II	III		I	II	III
J <sub>1</sub>	26	52	80	(I×J) <sub>1</sub>	23	49	80
J <sub>2</sub>	35	60	80	(I×J) <sub>2</sub>	38	63	80
计	61	112	160	计	61	112	160

各因素的效应为:

$$CF = \frac{61^2}{160}W_1 - \frac{112^2}{160}W_2 = 471.9$$

$$S_A = \frac{(29-33)^2}{160} W_1 + \frac{(52-60)^2}{160} W_2 = 2.1$$

$$S_B = \frac{(33-28)^2}{160} W_1 + \frac{(56-56)^2}{160} W_2 = 0.9$$

$$S_C = 3.2, S_D = 0.4, S_E = 1.1, S_F = 1.8, S_G = 0.0$$

$$S_H = 15.7, S_{C \times H} = 1.3, S_I = 20.4, S_J = 4.0, S_{I \times J} =$$

11.8

总波动  $S_T = 320$

一次误差波动的  $S_{e_1}$  可以通过把误差对应的 5、9、15 列的  $S$  值相加求得数据如表 46.3-46 所示。

表 46.3-46

	I	II	III
(5) <sub>1</sub>	30	55	80
(5) <sub>2</sub>	31	57	80
计	61	112	160
	I	II	III
(9) <sub>1</sub>	32	54	80
(9) <sub>2</sub>	29	58	80
计	61	112	160
	I	II	III
(15) <sub>1</sub>	30	56	80
(15) <sub>2</sub>	31	56	80
计	61	112	160

$$S_{e_1} = \frac{(30-31)^2 + (32-29)^2 + (30-31)^2}{160} W_1 + \frac{(55-57)^2 + (54-58)^2 + (56-56)^2}{160} W_2 = 0.8$$

二次误差

$$S_{e_2} = S_T - (S_A + S_B + \dots + S_{e_1})$$

$$= 320 - (2.1 + 0.6 + \dots + 0.8) = 256.6$$

因而得方差分析表(见表 46.3-47)。

表 46.3-47 积累法的方差分析表

因素	$f$	$S$	$V$	$F$	$\rho(\%)$
A	2	2.1	1.05	—	—
B	2	0.6	0.30	—	—
C	2	3.2	1.60	1.8	—
D	2	0.4	0.20	—	—
E	2	1.1	0.55	—	—
F	2	1.8	0.90	—	—
G	2	0.0	0.00	—	—
H	2	15.7	7.85	8.9	4.4
G×H	2	1.3	0.65	—	—
I	2	20.4	10.20	11.4	5.8
J	2	4.1	2.05	2.3	—
I×J	2	11.8	5.90	6.6	3.1
$e_1$	5	0.8	0.13	—	—
$e_2$	288	256.6	0.891	—	—
计	318	320.0			

自由度 6 的  $e_1$ , 无显著意义。B、D、G 等比  $e_1$  小, 所以可以把一次误差视为很小, 但无必要与  $e_2$  合并。

### 6.3.2 累计法、因素效应估计、最宜条件的决定

根据辅助表作 H 的主效应表与 I 与 J 的二元表,

如表 46.3-48 所示。

表 46.3-48 H 主效应与 I、J 二元表

	G	T	B	计
H <sub>1</sub>	42.50	11.25	16.25	100
H <sub>2</sub>	33.75	22.50	43.75	100
平均	38.125	81.875	30.000	100
	G	T	B	计
I <sub>1</sub> J <sub>1</sub>	32.5	37.5	30.0	100
I <sub>1</sub> J <sub>2</sub>	62.5	35.0	2.5	100
I <sub>2</sub> J <sub>1</sub>	32.5	27.5	40.0	100
I <sub>2</sub> J <sub>2</sub>	25.0	27.5	47.5	100
平均	38.125	31.875	30.000	100

根据上表的数据可以得出最宜条件为 H<sub>1</sub>I<sub>1</sub>J<sub>2</sub>。在 H<sub>1</sub>I<sub>1</sub>J<sub>2</sub> 条件下, 估计工序平均值用欧米加法。查附表分贝值得

$$\hat{\mu}_1 = (-1.31) + (2.22) + (+1.09) = 2.00(\text{dB})$$

$$\mu_1 = 61.3\%$$

$$\mu_2 = 98.5\%$$

在 H<sub>1</sub>I<sub>1</sub>J<sub>2</sub> 条件下 G、T、B 的出现比例为 (61.3, 37.2, 1.5)。

### 6.4 说明

实验设计中所取的试样数明显过少, 至少应取 20 个, 尽可能多取, 如果可能应该制造 50 个试样。

## 7 电灯泡的瓦数, 功率和寿命实验

### 7.1 目的

实验目的为了找出灯泡制造工序对灯泡三大特性瓦数、效率、寿命 特别是它们的波动的效应, 制造出平均值适宜的, 并尽可能使这些特性值相协调的灯泡, 同时还对灯丝损失进行了分析。

### 7.2 因素与水平

实验选定了下列工序。

1) 绕灯丝工序 钨丝自动地以 300r/min 的速度, 绕铁心轴打卷, 制成灯丝。不同标准生产的灯泡瓦数、效率、寿命平均值和波动都不同, 以美国制与英国制为例, 把心轴的供应者取为因素 B—B<sub>1</sub>: 英国制心轴, B<sub>2</sub>: 美国制心轴。其次, 采纳技术人员的意见, 将心轴的拉力取为因素 A, 其水平:

A<sub>1</sub>—拉力强, A<sub>2</sub>—拉力弱, A<sub>3</sub>—拉力中等

2) 灯丝退火工序 绕心轴卷成的钨丝, 连续通过小型隧道式电炉进行退火。温度与通过速度取为因素 D、C。C 水平取值:

C<sub>1</sub>: 标准速度, C<sub>2</sub>: 速度减半

对应于 C<sub>1</sub>, 温度水平取下列数值:

$D_1 D_2 D_3$

$C_1 750 765 735$  (735~765℃的三等分)

同样对应于  $C_2$ , 温度水平取下列数值:

$D_1 D_2 D_3$

$C_2 665 680 650$  (650~680℃的三等分)

根据温度表指示偏差±5℃。

3) 灯丝切断工序 为把退火后的灯丝切成所需的长度, 截断机型号有两种, 取为因素  $E$

$E_1$ : 原来的截断机

$E_2$ : 新截断机

还有灯丝的长度波动会影响特性值的波动, 以及灯丝的损失。另外取灯丝的抗拉强度为  $F$ 。

$F_1$ : 抗拉力强

$F_2$ : 抗拉力弱

4) 瓦斯封入工序 灯丝除去心轴后, 放在心棒上端, 插入玻璃管中, 用密封机焊接后, 吸出空气, 封入瓦斯。选取三个因素:

瓦斯量  $G$ :  $G_1=0.85, G_2=0.82, G_3=0.80$  (气/瓦)

瓦斯混合比  $H$ :  $H=86\%, H=92\%$

烘箱温度  $I$ :  $I_1=400, I_2=420, I_3=440$  (°C)

### 7.3 配列与数据

选用  $L_{27}(3^{13})$  正交表。全部二水平因素都加上虚拟水平。第一水平与第二水平重要的加倍进行实验。由于因素的水平与世纪的水平是随机对应的。

最重要因素为  $C$ , 其次为  $D, G$ , 所以把  $C \times D, D \times G$  配列于  $L_{27}(3^{13})$  正交表中, 以便分析。关于各次实验制造几个灯泡的问题, 重点分析波动的制造 10 个左右。在做几种灯泡的同型实验时, 各次实验只调查 4 个灯泡。配列与数据如表 46.3-49 所示。

表 46.3-49 配列与数据

	3	9	1	2	10	11	5	12	13	瓦数	
	A	B	C	E	E	F	G	H	I	平均	R
1	1	1	1	1	1	1	1	1	1	39.24	0.77
2	2	2	1	1	2	2	2	1	2	39.62	0.46
3	3	2	1	1	1	1	3	2	3	40.14	0.00
4	2	2	1	2	2	1	1	2	3	39.47	0.46
5	3	2	1	2	1	1	2	1	1	40.02	0.86
6	1	1	1	2	1	2	3	1	2	38.93	0.23
7	3	2	1	3	1	2	1	1	2	40.08	0.23
8	1	1	1	3	1	1	2	2	3	39.07	0.81
9	2	2	1	3	2	1	3	1	1	40.02	0.46
10	1	2	2	1	1	1	1	1	3	39.65	0.35
11	2	2	2	1	1	2	2	2	1	39.33	0.46
12	3	1	2	1	2	1	3	1	2	39.62	0.12
13	2	2	2	2	1	1	1	1	2	39.27	0.92
14	3	1	2	2	2	1	2	1	3	39.53	0.12
15	1	2	2	2	1	2	3	2	1	39.33	0.46
16	3	1	2	3	2	2	1	2	1	38.93	0.23
17	1	2	2	3	1	1	2	1	2	39.85	0.46
18	2	2	2	3	1	1	3	1	3	39.13	0.81

(续)

	3	9	1	2	10	11	5	12	13	瓦数	
	A	B	C	E	E	F	G	H	I	平均	R
19	1	2	1	1	2	1	1	2	2	40.02	0.46
20	2	1	1	1	1	2	2	1	3	37.92	1.04
21	3	2	1	1	1	1	3	1	1	39.82	0.58
22	2	1	1	2	1	1	1	1	1	39.19	0.58
23	3	2	1	2	1	1	2	2	2	40.34	0.58
24	1	2	1	2	2	2	3	1	3	39.04	0.23
25	3	2	1	3	1	2	1	1	3	39.77	0.58
26	1	2	1	3	2	1	2	1	1	39.76	0.57
27	2	1	1	3	1	1	3	2	2	39.04	0.45

(C×D)-三、四列 (C×G)-六、七列 e 根据虚拟水平

### 7.4 数据分析

数据等于瓦数的平均减去 40 之后乘以 100。拟水平采取后来加上的方法。计算使用下述的简易算法: 首先对一、二列组合求和, 三、四列根据其 9 个数据之和合计; 对一、五列组合求和; 六、七列按三、四列方法求和。一、八列组合, 一、十一列组合取四次和。

计算平均值与  $R$ , 需用辅助表。平均值  $m$  的辅助表如表 46.3-50 所示,  $R$  的辅助表如表 46.3-51 所示。

表 46.3-50  $m$  的辅助表

$k=9$	计	$k=3$	$C_1$	$C_1'$	$C_2$	计
$A_1$	-511	$D_1$	-100	-224	-140	-464
$A_2$	-701		-324			
$A_3$	-175					
$B_1$	853	$D_2$	-158	-143	-187	-488
$B_2$	-213		-301			
$B_3$	-321					
$E_1$	-510	$D_3$	-83	-143	-209	-435
$E_1'$	-478		-226			
$E_2$	-399					
$F_1$	-301	$G_1$	-121	-102	-215	-438
$F_1'$	-381		-223			
$F_2$	-705					
$H_1$	-525	$G_2$	-129	-198	-129	-456
$H_1'$	-429		-327			
$H_2$	-433					
$I_1$	-436	$G_3$	-91	-210	-192	-493
$I_2$	-323		-301			
$I_3$	-628					
计	-1387	计	-341	-510	-536	-1387
			-851			

计算平均值  $m$  的波动, 得

$$CF = \frac{(-1387)^2}{27} = 71251 \quad (46.3-1)$$

$$S_{T_1} = (-76)^2 + (-38)^2 + \dots + (-96)^2 - CF = 68418$$

$$S_A = \frac{1}{9} [(-511)^2 + (-701)^2 + (-175)^2] - CF = 15765$$



表 46.3-51 R 的辅助表

k=9	计	k=3	C <sub>1</sub>	C <sub>1</sub> '	C <sub>2</sub>	计
A <sub>1</sub>	434	D <sub>1</sub>	123	208	93	424
A <sub>2</sub>	565		331			
A <sub>3</sub>	330					
B <sub>1</sub>	436	D <sub>2</sub>	155	139	150	444
B <sub>2</sub>	439		294			
B <sub>3</sub>	454					
E <sub>1</sub>	574	D <sub>3</sub>	150	161	150	161
E <sub>1</sub> '	444		311			
E <sub>2</sub>	311					
F <sub>1</sub>	487	G <sub>1</sub>	146	162	150	458
F <sub>1</sub> '	450		308			
F <sub>2</sub>	392					
H <sub>1</sub>	544	G <sub>2</sub>	213	219	104	536
H <sub>1</sub> '	393		432			
H <sub>2</sub>	392					
I <sub>1</sub>	497	G <sub>3</sub>	69	127	139	335
I <sub>2</sub>	392		169			
I <sub>3</sub>	440					
计	1329	计	428	508	393	1329
			936			

$$S_B = \frac{[-853 \times 2 - (-534)]^2}{9 \times 2^2 + 18 \times (-1)^2} = \frac{(-1172)^2}{54} = 25437$$

$$S_C = \frac{[-851 - 2 \times (-536)]^2}{18 \times 1^2 + 9 \times (-2)^2} = 904$$

S<sub>D</sub> = 156

S<sub>E</sub> = 669

S<sub>F</sub> = 9851

S<sub>G</sub> = 147

S<sub>H</sub> = 143

S<sub>I</sub> = 5283

$$S_{C \times D} = \frac{1}{6} [(-324)^2 + (-301)^2 + (-226)^2] + \frac{1}{3} [(-140)^2 + (-187)^2 + (-200)^2] - CF - S_C - S_D = 1548$$

S<sub>C × C</sub> = 2124

S<sub>e</sub> = S<sub>T</sub> - (S<sub>A</sub> + S<sub>B</sub> + ... + S<sub>C × C</sub>) = 6400

$$V_{e_2} = \left( \frac{\bar{R}}{d_2} \right)^2 = \left( \frac{1}{2.059} \times \frac{1329}{27} \right)^2 = 571$$

准确的自由度为 3 × 27 × 效率系数, 但是取 81 也可以, m 的方差分析表为表 46.3-52。

表 46.3-52 m 的方差分析表

因素	f	S	V
A	2	15765	7882
B	1	25437	25437
C	1	904	904
D	2	156	78
E	1	669	669
F	1	9815	9815
G	2	174	87
H	1	143	143
I	2	5283	2642
C × D	2	1548	774

(续)

因素	f	S	V
C × G	2	2124	1062
e	9	6400	711
e <sub>1</sub> (根据极差)	81	46291	571
	107	114709	

R 的方差分析可以按通常的测量值考虑, 实际上取对数更好。

CF = 65406

S<sub>T</sub> = 77<sup>2</sup> + 46<sup>2</sup> + ... + 46<sup>2</sup> - CF = 17573

S<sub>A</sub> =  $\frac{1}{9} [(434)^2 + (565)^2 + (330)^2] - CF = 6400$

⋮

S<sub>e</sub>, S<sub>T</sub> (S<sub>A</sub> + ... + S<sub>C × G</sub>) = 3814

R 的方差分析表如表 46.3-53 所示。此结果表明只有 E 对 R 有显著意义。

表 46.3-53 R 的方差分析

因素	f	S	V
A	2	3082	1541
B	1	8	8
C	1	417	417
D	2	77	38
E	1	3904	2904
F	1	434	434
G	2	2282	1141
H	1	434	434
I	2	614	307
C × D	2	759	380
C × G	2	2748	1374
e	9	3814	424
计	26	18573	

### 7.5 估计

对 m 来说, I 的效果影响率小而忽略。

$$A_1 = 40 + \frac{-511}{9} \times \frac{1}{100} = 39.43$$

$$A_2 = 40 + \frac{-701}{9} \times \frac{1}{100} = 39.22$$

$$A_3 = 40 + \frac{-175}{9} \times \frac{1}{100} = 39.81$$

$$B_1 = 40 + \frac{-853}{9} \times \frac{1}{100} = 39.05$$

$$B_2 = 40 + \frac{-534}{18} \times \frac{1}{100} = 39.70$$

$$F_1 = 40 + \frac{-682}{18} \times \frac{1}{100} = 39.62$$

$$F_2 = 40 + \frac{-705}{9} \times \frac{1}{100} = 39.22$$

把上述以外的因素全部包括在内, 则有

S<sub>e</sub> = 63692, f<sub>2</sub> = 103, V<sub>e</sub> = 618

所以对应于  $n_c=9$  的置信界限为

$$\pm \sqrt{3.94 \times 618 \times \frac{1}{9} \times \frac{1}{100}} = \pm 0.16$$

对应于  $n_c=18$  的置信界限为  $1/\sqrt{2}$ , 为  $\pm 0.12$

对于  $R$

$$E_1 = \frac{1018}{18} \times \frac{1}{100} = 0.57 \pm 0.12$$

$$E_2 = \frac{311}{9} \times \frac{1}{100} = 0.35 \pm 0.17$$

包括其他因素的  $S_e=14669$ , 自由度为 25, 所以置信界限为  $+0.12$  与  $\pm 0.17$ 。

平均值  $m$  的因素效应分析、波动的因素效应分析如图 46.3-6、46.3-7 所示。

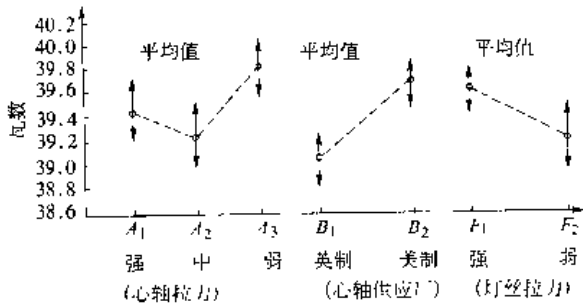


图 46.3-6 平均值的因素效应

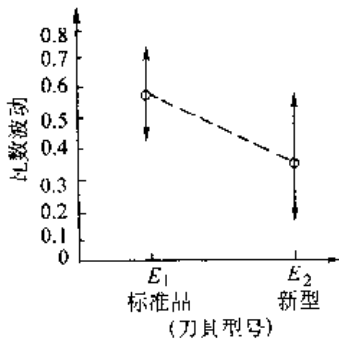


图 46.3-7 波动的因素分析

### 7.6 结论

因为只列出瓦数数据, 就瓦数结论进行简单分析:

心轴的抗拉力变弱瓦数平均值比其他情况大 0.4~0.5W。截断机的抗拉力越好, 瓦数值越小。本实验的目的是分析波动。影响波动即灯泡瓦数均匀性的因素只有截断机一种, 一般认为新型的截断机可以使产品波动减少 40%。应该改用新型的截断机。退火工序和瓦斯封入工序对瓦数值几乎没有什么影响, 但是对灯泡的寿命与效率有很大的影响。所以采用生产率高的退火。

## 8 氧化膜实验

### 8.1 目的

用铝合金制造的飞机零件, 为改善其耐腐蚀性、硬度、色调等, 必须在表面形成氧化膜。形成氧化膜的工序为电解氧化工序, 某公司常出现阳极膜厚度不均匀、有气泡的缺陷。检验方法是, 在零件表面加上 50V 的电压, 形成足够的阳极膜厚度就为合格, 如果厚度偏薄或者有气泡, 通过微型电灯泡来显示阳极膜厚度不够的位置。不良率高达 20%, 虽然经过返修可以成为正品, 但从生产率生产成本考虑, 就不经济。而且有时容易误检, 将不良品转入装配工序。

因而通过工序作业的改进, 以改善膜厚的不良性。为解决问题公司提出将电解槽的铬酸浓度有 20~50g/L 变为 50~60g/L 的建议。

### 8.2 因素与水平

阳极膜的形四个工序:

1) 清洗工序 为除去飞机用铝合金零件和铝板表面的油污等, 首先, 在三氯化化乙烯蒸汽中掉挂 8min 左右; 其次在碱溶液中浸泡数分钟。由于水洗, 选在在二氯化化乙烯蒸汽中掉挂时间为因素。水平为现行的 8min, 10min, 12min。

因素 A:  $A_1=8\text{min}, A_2=10\text{min}, A_3=12\text{min}$

2) 电解氧化工序 洗净的零件为形成阳极氧化膜, 作为与纯铝线连接的阳极挂在铬酸溶液中。通电流在 10min 内把电压慢慢地从 0V 到 40V, 在 40V 保持 30min, 然后在 5min 内慢慢地提高到 50V, 保持 5min, 切断电流。这是电解氧化工序的作业标准。

为求最佳操作曲线, 需弄清以下六个问题: 开始用几分钟(因素 A)把电压提高到多少伏(因素 B), 在此电压下保持几分钟(因素 C), 其次用几分钟(因素 E)把电压提高到多少伏(因素 D), 在该电压下保持几分钟(因素 F)。作为因素只能提出下述两种: 在短时间高电压; 长时间低电压。

关于电压, 最初提高的伏值到最后提高到的伏值的两问题, 所以电压因素的水平:

$D_1: 38\sim 50\text{V}, D_2: 40\sim 50\text{V}, D_3: 45\sim 52\text{V}$

$D_1$ : 取为 38V, 与现行标准有些少许改变, 因为预想把电压改低可能有使零件质量变坏的危险。

关于时间, 在开始把电压提高到一定水平后, 要保持较长时间(按标准为 30min), 故以保持时间为重点选定水平。但是电压取低时, 时间应该长一些。时间因素 E 的水平根据 D 的水平确定。(见表 46.3-54)。

表 46.3-54 时间周期

电压/V	$E_1$	$E_2$	$E_3$
$D_1$ 38~50	10-30 5 5 (50)	10-35-5-5 (55)	10-40-5-3 (60)
$D_2$ 40~50	10-25-5-5 (45)	10 30 5 5 (50)	10-35-5-5 (55)
$D_3$ 45~52	10-20-4-4 (38)	10-25-4 4 (48)	10 30 -4-4 (48)

以  $DE$  为例说明,含义是:在开始的 10min,电压提高到 38V,在 38V 保持 30min,然后在 5min 内慢慢提高到 50V,在 50V 保持 5min,合计 50min。

其次,铬酸的浓度与温度。在浓度方面,某公司的建议是 25~50g/L 质量管理部门认为浓一些好,所以水平可选择: $c=50\text{g/L}$ , $c=40\text{g/L}$ , $c=30\text{g/L}$ 。变浓较容易,稀释困难些。

在铬酸溶液温度方面,取其范围的最大值与最小值,选定水平如下: $F_1=100^\circ\text{F}$ , $F_2=103^\circ\text{F}$ , $F_3=106^\circ\text{F}$ ,温度控制加温通过高温蒸汽,降温通过时间冷却或加入大量冷液,这因素水平也较难改变。

3) 水洗时间 完成工序作业的零件,投入水中洗掉表面附着的铬酸溶液,该时间水平选:

$$G_1=2\text{min}, G_2=3\text{min}$$

4) 热水浸泡时间 把零件浸入  $150^\circ\text{F}$  的热水中,以增加阳极膜的韧性,浸泡时间取因素  $H$ :

$$H_1=3\text{min}, H_2=4\text{min}, H_3=5\text{min},$$

5) 表示因素 以上为控制因素。在这里把零件的形状作为标示因素提出。大型的形状复杂的零件较难形成均匀的阳极膜,因此取零件形状为标示因素  $B$ 。

该因素水平如下:

$B_1$ =中型零件, $B_2$ =小型零件

$B_3$ =大型零件,形状复杂零件

有关因素重要性,质量管理部门认为浓度因素  $C$  与零件形状因素  $B$  最为重要。因而两者的交互作用尽可能不要与其他效应混杂。

### 8.3 配列与实验

由于时间短,同时为了因素  $C, D, E, F$  的实验顺序随机化程度小于其他因素,在  $L_{27}(3^{13})$  正交表中,使  $C, F$  对应于一列与二列,使  $D, E$  二因素对应于三列、四列。即把因素  $C, D, E, F$  视为一次因素。再者把因素  $B$  对应于五列,为了把  $C \times B$  的交互作用同其他因素分离,未把其他因素同对应于  $1 \times 5$  的六、七列对应。

因素  $G$  取二水平,加上一个虚拟水平,把第一、二水平对应于  $G_1$ ,第三水平对应于  $G_2$ 。 $A, G, H$  同一至七列以外的列对应。再者,由于把各因素水平的表示值与正交表字母 1、2、3 的对应随机化,所以下列的配列不一定符合自然顺序。

配列表如表 46.3-55。

表 46.3-55 配列

No.	因素列							
	C	D	E	F	B	A	G	H
1					2	3	1	3
2	3	2	1	1	1	1	1	2
3					3	2	2	1
4					2	1	1	1
5	3	1	2	2	1	2	2	3
6					3	3	1	2
7					2	2	2	2
8	3	3	3	3	1	3	1	1
9					3	1	1	3
10					2	3	1	1
11	2	3	2	1	1	1	2	3
12					3	2	1	2
13					2	1	2	2
14	2	2	3	2	1	2	1	1
15					3	3	1	3
16					2	2	1	3
17	2	1	1	3	1	3	1	2
18					3	1	2	1
19					2	3	2	2
20	1	1	3	1	1	1	1	1
21					3	2	1	3
22					2	1	1	3
23	1	3	1	2	1	2	1	2
24					3	3	2	1
25					2	2	1	1
26	1	2	2	3	1	3	2	2
27					3	1	1	3

因素  $C, D, E, F$  在每三次实验中都保持了同一条件,所以这些因素与进行九次实验是一样的。因而在实验号 1 到 27 中,例如 No. 1~No. 3 实验,在各次实验中是把大、中、小三种零件各供应 2 个,所以要取  $B_2$  (小)、 $B_1$  (中)、 $B_3$  (大) 的零件各两个,同时放入清洗槽中, $B_2$  洗 12min ( $A_1$ ),  $B_1$  洗 8min ( $A_1$ ),  $B_3$  洗 10min ( $A_2$ ) 后分别取出送往形成工序。把这六个零件投入到温度  $100^\circ\text{F}$  ( $B_1$ )、浓度调节到  $30\text{g/L}$  ( $C_3$ ) 的槽中,按  $D_2E_1$  的水平组合,即在 10min 内把电压慢慢提高到 40V,保持 25min,然后用 5min 把电压升到 50V,保持 5min 之后取出。从阳极膜形成槽取出后放入水中, $B_1, B_2$  洗 2min 取出, $B_3$  洗 3min 取出。最后同时放入热水中, $B_2$  浸泡 5min 取出, $B_1$  浸泡 3min 取出, $B_3$  浸泡 4min 取出。干燥后各零件上适当地选六个部位,用静电容量计测量微法拉值。

### 8.4 数据

曾用铝板切取小试验片,观测阳极膜形成条件的优劣,为观察方法是否妥当,除大中小零件各取 2 个外,还加一个试验片。该试验片也测量 6 个值。数据中的“试”表示试验片的值,1、2 是零件本身的值。得数据表 46.3-56。

表 46.3-56 静电容量试验数据

№1(小)			№8(中)			№15(大)		
试	1	2	试	1	2	试	1	2
0.4	0.4	∞	1.0	1.0	1.0	1.5	4.0	1.0
∞	∞	∞	1.0	1.0	1.0	2.0	∞	1.5
∞	∞	∞	1.0	1.0	1.0	∞	6.0	1.5
∞	∞	∞	∞	∞	1.0	1.5	∞	1.0
0.5	∞	∞	1.0	1.0	2.0	2.5	6.0	2.0
∞	∞	0.2	1.0	1.0	1.0	∞	∞	1.5
№2(中)			№9(大)			№16(小)		
试	1	2	试	1	2	试	1	2
∞	9.0	∞	1.0	1.5	1.5	1.7	0.3	1.0
∞	∞	∞	1.5	2.0	2.0	2.0	0.5	1.0
∞	∞	∞	1.0	1.0	1.0	2.0	0.5	0.8
∞	∞	∞	1.0	2.0	2.0	2.5	∞	1.0
∞	∞	∞	1.5	2.0	1.0	2.8	0.5	1.0
∞	∞	∞	1.0	1.0	1.0	2.5	0.8	0.5
№3(大)			№10(小)			№17(中)		
试	1	2	试	1	2	试	1	2
1.0	∞	∞	1.0	∞	0.5	2.0	0.5	0.5
11.0	∞	1.0	∞	0.5	∞	2.5	0.5	1.0
∞	∞	∞	∞	1.0	∞	∞	0.5	1.0
2	∞	∞	∞	∞	∞	2.0	0.5	0.7
3	∞	∞	∞	0.5	∞	2.5	0.5	0.5
∞	∞	∞	∞	∞	∞	∞	1.0	0.5
№4(小)			№11(中)			№18(大)		
试	1	2	试	1	2	试	1	2
1.0	1.0	1.0	1.0	0.5	0.5	2.0	6.0	2.0
∞	∞	1.0	∞	0.5	0.5	2.5	7.0	2.0
1.0	1.0	1.0	∞	∞	∞	2.0	5.0	2.5
∞	1.0	3.0	∞	∞	∞	2.0	5.0	2.5
1.0	3.0	2.0	∞	0.5	∞	2.5	∞	2.0
1.0	2.0	1.0	∞	∞	∞	2.0	4.0	2.5
№5(中)			№12(大)			№19(小)		
试	1	2	试	1	2	试	1	2
1.0	2.0	1.0	∞	1.0	0.6	1.5	0.8	0.6
1.0	2.0	1.0	1.5	∞	0.6	1.8	0.7	0.9
1.0	1.0	2.0	∞	∞	∞	1.8	1.0	1.0
∞	3.0	1.0	1.0	1.0	1.0	2.0	1.2	1.0
1.0	2.0	2.0	∞	1.0	∞	2.0	0.8	0.8
1.0	1.0	2.0	∞	∞	∞	2.0	1.0	0.8
№6(大)			№13(小)			№20(中)		
试	1	2	试	1	2	试	1	2
1.5	∞	1.0	0.5	1.0	0.5	2.0	0.5	0.5
∞	∞	2.0	1.0	1.0	0.7	1.8	0.6	0.3
1.5	1.0	2.0	1.5	1.0	1.0	2.3	0.5	0.6
1.5	1.0	5.0	1.5	∞	0.5	2.0	0.5	0.5
1.5	1.0	2.0	1.5	1.0	1.0	2.0	0.6	0.5
∞	1.0	2.0	1.0	1.0	1.5	2.3	0.3	0.3
№7(小)			№14(中)			№21(大)		
试	1	2	试	1	2	试	1	2
1.0	1.0	2.0	0.8	1.5	1.5	2.5	∞	1.5
1.0	1.0	1.5	1.0	1.0	1.5	3.0	8.0	1.5
1.0	1.0	2.0	1.0	1.0	1.0	3.0	∞	2.0
∞	∞	1.5	∞	∞	1.5	3.0	∞	2.0
1.0	1.0	1.0	1.5	1.5	1.5	2.5	9.0	∞
1.0	1.0	1.0	1.5	1.5	1.0	3.0	∞	1.5

(续)

№22(小)			№24(大)			№26(中)		
试	1	2	试	1	2	试	1	2
2.5	1.0	0.5	0.5	9	2.5	0.7	0.2	0.3
2.5	1.0	1.0	1.0	10	3.0	0.7	0.4	0.4
3.0	0.5	1.0	1.8	15	3.0	0.5	∞	∞
2.5	1.0	1.2	1.5	20	3.5	1.0	0.2	0.3
3.0	1.0	0.5	1.0	∞	2.5	0.7	∞	∞
2.7	0.8	0.7	1.5	20	2.5	0.5	0.2	0.4
№23(中)			№25(小)			№27(大)		
试	1	2	试	1	2	试	1	2
2.0	1.0	1.5	0.5	0.2	0.2	0.5	1.0	0.4
3.0	2.0	1.5	0.5	0.3	0.3	0.7	∞	0.5
2.5	1.0	1.5	0.5	0.2	0.3	∞	0.6	0.5
2.0	1.5	1.7	0.7	0.5	∞	0.5	∞	0.4
3.0	1.5	1.7	0.5	0.3	0.2	0.7	0.6	0.3
2.0	2.0	2.0	0.5	0.3	0.3	0.5	∞	0.4

再者按照一定的方法通入 50V 电压时,还收集了电路中的微型灯泡是否亮的所谓通电实验数据。在八个部位进行实验以查合格或不合格。

表 46.3-57 是就各个零件的八个部位做通电实验,因膜厚不够使电灯泡亮的不合格次数。

表 46.3-57 通电试验数据

№	大			中			小		
	试	1	2	试	1	2	试	1	2
1	2	3	3	5	2	3	4	2	4
2	0	0	5	1	1	1	0	1	2
3	0	0	0	0	0	0	0	2	0
4	5	0	5	5	4	2	6	4	5
5	2	0	2	2	0	2	3	1	1
6	2	1	0	0	0	2	1	2	1
7	1	0	1	0	0	0	1	0	0
8	4	2	5	3	2	1	2	2	2
9	0	0	0	0	0	1	0	1	1

作辅助表

把静电容量实验的数据测量值分为四组,使用计数分类值的分析法。

I'——测量值为 0~1.5 的一组。该组零件的膜厚较薄,经济性不好,但质量完好。

II'——测量值为 1.5~3.0 的一组。膜厚足够,最为理想。

III'——测量值超过 3.0 的一组(有限值)。这组零件膜厚过薄,所以希望尽量减少。

IV'——测量值无限大的一组。这组零件主要气泡等缺陷,希望完全消除。

各种因素的效应与一般零件有差异,所以只对实际零件进行分析。因而只就实际零件来计数,各个测量值落入 I'、II'、III'、IV' 各组内的频数,如表 46.3-58 所示。

表 46.3-58 配列与分类值数据

实验号	C	D	E	F	B	A	G	H	零件 1				零件 2				计			
									I'	II'	III'	IV'	I'	II'	III'	IV'	I'	II'	III'	IV'
1	3	2	1	1	2	3	1	3	1	0	0	5	1	0	0	5	2	0	0	10
2	3	2	1	1	1	1	1	2	0	0	1	5	0	0	0	6	0	0	1	11
3	3	2	1	1	3	2	2	1	0	0	0	6	0	0	0	6	0	0	0	12
4	3	1	2	2	2	1	1	1	3	2	0	1	3	3	0	0	6	5	0	1
5	3	1	2	2	1	2	2	3	3	3	0	0	2	4	0	0	5	7	0	0

(续)

实验号	C	D	E	F	B	A	G	H	零件 1				零件 2				计				
									I'	II'	III'	IV'	I'	II'	III'	IV'	I'	II'	III'	IV'	
6	3	1	2	2	3	3	1	2	4	0	0	2	1	4	1	0	5	4	1	2	
7	3	3	3	3	2	2	2	2	5	0	0	1	4	2	0	0	9	2	0	1	
8	3	3	3	3	1	3	1	1	5	0	0	1	4	2	0	0	9	2	0	1	
9	3	3	3	3	3	1	1	3	3	3	0	0	4	2	0	0	7	5	0	0	
10	2	3	2	1	2	3	1	1	3	0	0	3	1	0	0	5	4	0	0		
11	2	3	2	1	1	1	2	3	3	0	0	3	2	0	0	4	5	0	0		
12	2	3	2	1	3	2	1	2	3	0	0	3	3	0	0	3	6	0	0		
13	2	2	3	2	2	1	2	2	5	0	0	1	6	0	0	0	11	0	0		
14	2	2	3	2	1	2	1	1	5	0	0	1	6	0	0	0	11	0	0		
15	2	2	3	2	3	3	1	3	0	0	3	3	5	1	0	0	5	1	1		
16	2	1	1	3	2	2	1	3	5	0	0	1	6	0	0	0	11	0	0		
17	2	1	1	3	1	3	1	2	6	0	0	0	6	0	0	0	12	0	0		
18	2	1	1	3	3	1	2	1	0	0	5	1	0	6	0	1	0	6	5		
19	1	1	3	1	2	3	2	2	6	0	0	0	6	0	0	0	12	0	0		
20	1	1	3	1	1	1	1	1	6	0	0	0	6	0	0	0	12	0	0		
21	1	1	3	1	3	2	1	3	0	0	2	4	3	2	0	1	3	2	2		
22	1	2	1	2	2	1	1	3	6	0	0	0	6	0	0	0	12	0	0		
23	1	2	1	2	1	2	1	2	4	2	0	0	3	3	0	0	7	5	0		
24	1	2	1	2	3	3	2	1	0	0	5	1	0	5	1	0	5	5	6		
25	1	3	2	3	2	2	1	1	6	0	0	0	5	0	0	1	11	0	0		
26	1	3	2	3	1	3	2	3	4	0	0	2	4	0	0	2	8	0	0		
27	1	3	2	3	3	1	1	2	3	0	0	3	6	0	0	0	9	0	0		
计	1	4	3	2	5	8	9	13									182	44	18	80	324

8.5 累积法分析

作累积频数的辅助表如表 46.3-59、表 46.3-60 所示。

表 46.3-59 累积频数的辅助表

	I'	II'	III'	IV'		I'	II'	III'	IV'
A <sub>1</sub>	62	78	84	108	B <sub>1</sub>	69	83	84	108
A <sub>2</sub>	63	79	81	108	B <sub>2</sub>	78	85	85	108
A <sub>3</sub>	57	69	79	108	B <sub>3</sub>	35	58	75	108
计	182	226	244	324	计	182	226	244	324
C <sub>1</sub>	74	86	94	108	E <sub>1</sub>	44	60	72	108
C <sub>2</sub>	65	72	80	108	E <sub>2</sub>	59	75	76	108
C <sub>3</sub>	43	68	70	108	E <sub>3</sub>	79	91	96	108
计	182	226	244	324	计	182	226	244	324
D <sub>1</sub>	66	90	98	108	F <sub>1</sub>	44	46	49	108
D <sub>2</sub>	57	58	62	108	F <sub>2</sub>	62	89	99	108
D <sub>3</sub>	59	78	84	108	F <sub>3</sub>	76	91	96	108
计	182	226	244	324	计	182	226	244	324
H <sub>1</sub>	53	71	82	108	G <sub>1</sub>	132	156	163	216
H <sub>2</sub>	71	82	84	108	G <sub>2</sub>	50	70	81	108
H <sub>3</sub>	58	73	78	108					
计	182	226	244	324	计	182	226	244	324

表 46.3-60 BC 的二元表

		I'	II'	III'	IV'
C <sub>1</sub>	B <sub>1</sub>	27	32	32	36
	B <sub>2</sub>	35	35	35	36
	B <sub>3</sub>	12	19	27	36
C <sub>2</sub>	B <sub>1</sub>	28	28	28	36
	B <sub>2</sub>	26	26	26	36
	B <sub>3</sub>	11	18	26	36
C <sub>3</sub>	B <sub>1</sub>	14	23	24	36
	B <sub>2</sub>	17	24	24	36
	B <sub>3</sub>	12	21	22	36
计		182	226	244	324

累积法的计算除掉 IV 组, 求 I、II、III 组的权数 W<sub>1</sub>、W<sub>2</sub>、W<sub>3</sub>。

$$W_1 = \frac{324^2}{182 \times (324 - 182)} = 4.0619$$

$$W_2 = \frac{324^2}{226 \times (324 - 226)} = 4.7398$$

$$W_3 = \frac{324^2}{244 \times (324 - 244)} = 5.3779$$

因而

$$CF = \frac{182^2}{324} W_1 + \frac{226^2}{324} W_2 + \frac{244^2}{324} W_3 (f=3)$$

$$S_A = \frac{62^2 + 63^2 + 57^2}{108} W_1 +$$

$$\frac{78^2 + 79^2 + 69^2}{108} W_2 + \frac{84^2 + 81^2 + 79^2}{108} W_3 - CF$$

$$= 3.3 (f=6)$$

$$S_B = 60.8 (f=6)$$

$$S_C = 40.8 (f=6)$$

$$S_D = 57.6 (f=6)$$

$$S_E = 59.9 (f=6)$$

$$S_F = 153.6 (f=6)$$

$$S_H = 9.7 (f=6)$$

$$S_G = \left( \frac{132^2}{216} + \frac{50^2}{108} \right) \times W_1 + \left( \frac{156^2}{216} + \frac{70^2}{108} \right) \times W_2$$

$$+ \left( \frac{163^2}{216} + \frac{81^2}{108} \right) \times W_3 - CF$$

$$= 7.6 (f=6)$$

$$S_{B \times C} = \frac{27^2 + 35^2 + \dots + 12^2}{36} \times W_1$$

$$+ \frac{32^2 + 35^2 + \dots + 21^2}{36} \times W_2$$

$$+ \frac{32^2 + 35^2 + \dots + 22^2}{36} \times W_3$$

$$- CF \quad S_B - S_C = 23.6 (f=12)$$

$$S_{G_1} = 0$$

$$S_{e_2} = \frac{2^2+0^2+\dots+9^2}{12} \times W_1 + \frac{2^2+0^2+\dots+9^2}{12} \times W_2 + \frac{2^2+1^2+\dots+9^2}{12} \times W_3 - (CF+S_A+\dots+S_{B \times C})$$

$$= 42.3 \quad (f=21)$$

$$S_{e_3} = \frac{1^2+0^2+\dots+6^2}{6} \times W_1 + \frac{1^2+0^2+\dots+6^2}{6} \times W_2 + \frac{1^2+0^2+\dots+6^2}{6} \times W_3 - 2610.6$$

$$= 99.9 \quad (f=81)$$

$S_{e_1}$  是 9 次实验间的误差,  $S_{e_2}$  是 27 次实验间的误差,  $S_{e_3}$  是 54 个零件间的误差。

总波动用(测量值总数) × (I、II、III 的组数) 求出:

$$S_T = 324 \times 3 = 972 \quad (f=323 \times 3=969)$$

从  $S_T$  减去零件间波动(271.05 - 2151.4 = 559.1)

可得四次误差  $S_{e_4}$  同一零件的测量部位间的误差。

$$S_{e_4} = 972 - 559.1 = 412.9 \quad (f=5 \times 54 \times 3=810)$$

得到如表 46.3-61 所示的方差分析表。

表 46.3-61 方差分析表

因素	f	S	V	F
C(浓度)	6	40.8	6.80	4.89
D(电压)	6	57.6	9.60	6.91
E(时间)	6	59.9	9.98	7.18
F(温度)	6	153.6	25.60	18.42
$e_1$ (一次误差,槽间)	0	—	—	—
A(清洗时间)	6	3.3	0.55	—
B(大、中、小件间)	6	60.8	10.13	7.29
B×C	12	23.6	1.97	—
G(水洗时间)	3	7.6	2.53	—
H(热水浸泡时间)	6	9.7	1.62	—
$e_2$ (二次误差,实验间)	21	42.3	2.01	$e' = e_2 + e_1,$ $f=102,$ $V=1.39$
$e_3$ (三次误差,零件间)	81	99.9	1.23	—
$e_4$ (四次误差,测量部位间)	810	412.9	0.51	—
计	969	972.0	—	—

如果用  $e_4$  检验  $e_3$ , 则以 1% 的水平确认有显著意义。用  $e_3$  检验  $e_2$  无显著意义, 所以把  $e_2$  与  $e_3$  合并, 形成  $e'$ , 检验了二次因素。一次误差自由度为 0, 无情报。用  $e'$  作检验有显著意义是应该作大概有效应的结论。但应附上无统计保证。

## 9 电话交换机的颜色调整实验

### 9.1 目的

电话交换员在各自的服务台上, 面对着 400 个灯泡的点灭。进行着插头的插拔, 通话的应答与切断作业。实验目的是定量地调查以彩色为中心的作业环境, 对作业效率, 疲劳、交换员的身心有什么影响。作业效率的测度采用应答时间和切断时间。应答时间是从指示灯亮到插入插头开始应答的时间, 切断时间是通话

终了到切断的时间。疲劳以闪烁为中心进行调查, 色的喜好用调查卡进行调查。

### 9.2 因素与水平

所选的因素与水平

A: 色相 九水平。交换插头的面与电键盘的色, 用迈歇尔(Munsell)方式, 选取色相、亮度、彩度三个属性。

色相是在迈歇尔彩色标记表中, 取了九个系统(紫色未取)。如果十个系统全部选取时需要作部分追加。色相分为三个色相群。令其因素为这三个色相群:



B: 亮度。用迈歇尔值表示, 选取接近退色的 5、6、7 三水平:

$$B_1=5, B_2=6; B_3=7$$

C: 彩度, 彩度取反射率低的三水平。

$$C_1=1, C_2=2; C_3=3$$

D: 应答、切断灯泡的颜色。为表示因素, 使用白、红、绿, 为改善判别性, 混合使用这三种。

$$D_1=白, D_2=红; D_3=绿$$

E: 灯泡的辉度, 这与背景台面的颜色有密切关系。把灯泡的端子电压取为三水平。

$$E_1=11V, E_2=13V; E_3=16V$$

就白灯泡按上述三水平, 辉度变为  $0.014cd/cm^2$ ,  $0.042cd/cm^2$ ,  $0.15cd/cm^2$ 。

F: 照度, 背景的照明通过用 40W 的白色荧光灯适当地进行遮光, 形成三个水平。

$$F_1=明(电键盘 300 lx, 插头面 150 lx)$$

$$F_2=中(电键盘 150 lx, 插头面 75 lx)$$

$$F_3=暗(电键盘 80 lx, 插头面 40 lx)$$

G: 呼唤数(申请数), 为了了解繁忙程度对应答、切断时间或疲劳什么影响, 考虑实际作业数据之后, 取如下水平:

$$G_1=3.5, G_2=4.0, G_3=5.0 \quad (s)$$

按泊松输入, G 为平均时间间隔。

H: 交换员, 为了解担任实验员的状态, 交换员按技能分三等上、中、下三组。其中每一组编制每个人的明细表, 根据各个组中的相对顺序, 按工龄分类, 进而按年龄分类, 如此得 27 组。九个水平组合。各选出一个

交换员。此因素为标示因素。(见表 46.3-62)。

表 46.3-62

交换员	技能	工龄	年龄
$H_1$	上	短	中
$H_2$	上	中	大
$H_3$	上	长	小
$H_4$	中	长	大
$H_5$	中	短	小
$H_6$	中	中	中
$H_7$	下	中	小
$H_8$	下	长	中
$H_9$	下	短	大

Q: 区组因素, 九水平。

实验中随着时间的推移, 一方面熟练起来另一方面人员疲劳, 所以取实验顺序为因素。

按照实验日分为九水平:

$R_1$  = 第一日,  $R_2$  = 第二日, ...,  $R_9$  = 第九日

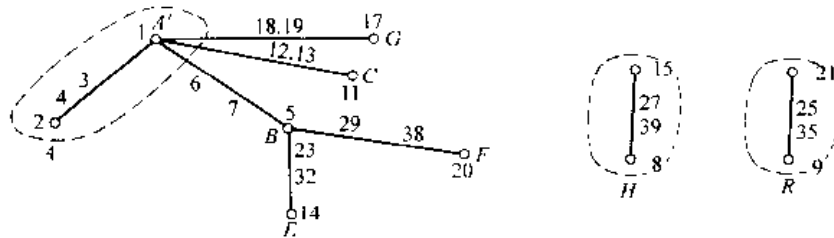


图 46.3-8 点线图

### 9.3 配列

由于把插头面、电键面的着色板做成了简单可变换的结构, 所以能一日作九种实验。使用了  $L_{81}$  正交表。为了尽量减少着色板的种类, 取  $A, B, C$  为一次因素, 列入 1~13。这样做 81 种着色板, 只准备 27 种就可以。把着色板减为 27 种, 与九种色相的交互作用就不能取全。因而重点考虑与色相群  $A'$  的交互作用。所考虑的交互作用取五种:  $A' \times B, A' \times C, A' \times G, A' \times E, A' \times F$ 。由于  $A, H, R$  有九水平, 所以不要忘记  $A, B, C$  是一次因素, 配列所需的点线图(见图 46.3-8)。利用  $L_{18}$  的点线图, 得到了图 46.3-8 的配列。灯泡色相  $D$  与应答切断的二水平一起, 与  $L_{81}$  进行直积:  $L_{81} \otimes L_3 \otimes L_2$ 。

配列如表 46.3-63 所示, 同时每个实验号的数百次呼唤随机进行, 用脉冲计数, 以 50ms 精度、区分灯泡的颜色、收集应答与切断所必须的时间, 然后脉冲去除, 求得平均应答时间。由于按每一应答切断收集数据, 所以数据量为  $81 \times 3 \times 2 = 486$  个。

表 46.3-63 配列结果一览表

因素 No.	彩 色				灯泡 亮度	实验 人码	呼 数	照 明	实 验 顺 序
	色相 群	色 相	亮 度	彩 度					
1	$A_1$	$2.5RP$	6	3	小	1	大	暗	1(8)
2						中	中	2(13)	
3						小	明	3(26)	
4				大		暗	4(36)		
5				中		中	5(45)		
6				小		明	6(51)		
7			大	暗	7(63)				
8			中	中	8(72)				
9			小	明	9(74)				
10			大	暗	4(47)				
11			中	中	5(29)				
12			小	明	6(40)				
13			大	暗	7(73)				
14			中	中	8(55)				
15			小	明	9(71)				
16			大	暗	1(19)				
17			中	中	2(1)				
18			小	明	3(70)				
19		大	暗	4(77)					
20		中	中	5(62)					
21		小	明	2(12)					
22		大	暗	3(21)					
23		中	中	1(4)					
24		小	明	5(40)					
25		大	暗	6(50)					
26		中	中	4(30)					
27		小	明	6(40)					

(续)

因素 No.	彩 色				灯泡 亮度	实验 人码	呼 数	照 明	实 验 顺 序
	色相 群	色 相	亮 度	彩 度					
28	$A_2$	$2.5Y$	6	3	小	2	大	暗	5(39)
29						中	中	6(48)	
30						小	明	4(31)	
31				大		暗	8(64)		
32				中		中	9(75)		
33				小		明	7(58)		
34			大	暗	2(15)				
35			中	中	3(23)				
36			小	明	1(5)				
37			大	暗	7(61)				
38			中	中	8(67)				
39			小	明	9(78)				
40			大	暗	1(6)				
41			中	中	2(10)				
42			小	明	3(24)				
43			大	暗	4(33)				
44			中	中	5(41)				
45			小	明	6(49)				
46	大	暗	3(22)						
47	中	中	1(3)						
48	小	明	2(11)						
49	大	暗	6(54)						
50	中	中	4(32)						
51	小	明	5(42)						
52	大	暗	9(79)						
53	中	中	7(56)						
54	小	明	8(66)						
55	大	暗	9(76)						
56	中	中	7(57)						
57	小	明	8(65)						
58	大	暗	3(20)						
59	中	中	1(7)						
60	小	明	2(17)						
61	大	暗	6(52)						
62	中	中	4(34)						
63	小	明	5(38)						

(续)

因素 No.	彩色			灯泡 亮度	实验 号码	呼 数	照 明	实验 顺序
	色 群	色 相	亮 度					
64				小	6	中	明	2(16)
65			6	中	4	小	暗	1(27)
66				大	5	大	中	3(2)
67				小	9	中	明	5(37)
68		2.5B	5	中	7	小	暗	6(53)
69				大	8	大	中	1(35)
70				小	3	中	明	8(67)
71			7	中	3	中	暗	9(80)
72				大	2	小	中	7(60)
73				小	1	大	中	1(28)
74			6	中	7	小	明	5(43)
75				大	5	大	暗	6(16)
76		2.5PB	5	中	3	小	明	7(59)
77				大	1	大	中	8(68)
78				小	2	中	暗	9(81)
79				中	6	小	明	1(9)
80			7	中	1	大	中	2(68)
81				大	5	中	暗	3(23)

DI以5%的水平有显著意义,但其影响率小,所以忽略。色相A<sub>1</sub>需要对其效应进行估计。与标示因素D、I的交互作用,由于只有采用平均的好水平,所以曲线图也只表示其结果(见图46.3-9)。

表 46.3 64 方差分析表

因素	f	S	V	ρ(%)
A(色相)	8	2067	258	0.4
B(亮度)	2	543	272	—
A×B	1	638	160	—
C(彩度)	2	1397	698	0.5
A×C	1	377	94	—
E(灯泡亮度)	2	29238	14619	11.9
B×E	1	96	24	—
F(照度)	2	3230	1615	1.2
B×F	1	671	168	—
G(呼唤数)	2	7590	3795	3.0
A×G	1	3448	862	1.2
H(交换员)	8	56969	7121	22.9
R(实验日)	8	2595	324	0.7
e <sub>1</sub> (一次误差)	26	1436	171	—
D(灯泡颜色)	2	30519	15260	12.4
DI(应答切断)	1	15015	15015	6.1
D×I	2	7136	3568	2.9
A×DI	10	3296	82	—
E×DI	10	11335	1154	4.3
G×DI	10	2568	257	—
H×DI	40	27060	675	9.3
E×B×DI	30	2769	92	—
t <sub>1</sub>	270	29806	114	—
t <sub>2</sub>	(394)	(45160)	(115)	23.2
计	485	242999		100.0

### 9.4 方差分析与估计

由于全部列出数据,数量过多,所以只列出方差分析的结果与因素效应的估计值。为了减少着色板的数量,一次因素A、B、C未区别一次误差与二次误差(见表16.3-64)。

因素D和I都是标示因素,故未加细分。还有G×

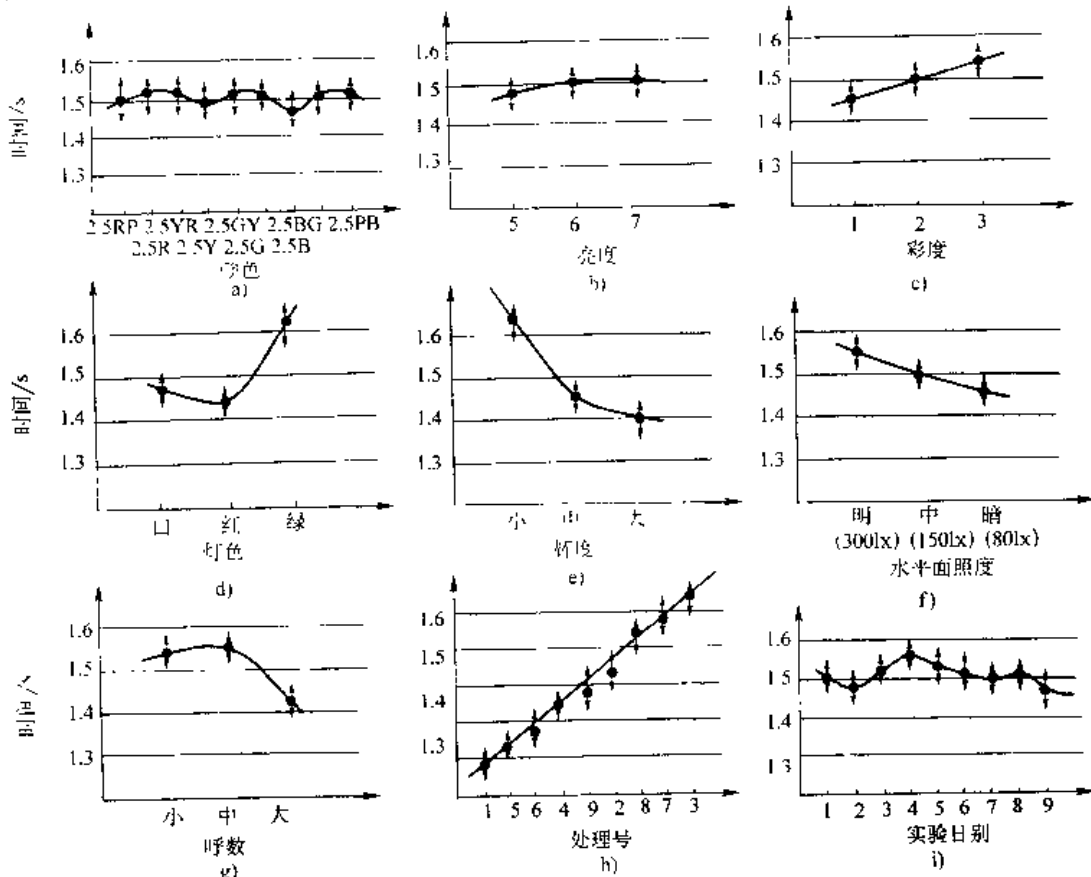


图 46.3 9 各种因素效应曲线



## 第 4 章 测量技术的实验设计方法

### 1 校正方法与测量误差

测量误差一般是测量值减去真值。

假设测量值为  $y$ , 此时的损失函数为  $L(y)$ 。假设真值为  $m$ , 则  $L(m)=0$ 。

以  $m$  为中心将损失函数  $L(y)$  进行泰勒展开, 得

$$\begin{aligned} L(y) &= L(m + y - m) \\ &= L(m) + \frac{L'(m)}{1!}(y - m) + \\ &\quad \frac{L''(m)}{2!}(y - m)^2 + \dots \end{aligned}$$

可知  $L(m)=0, L'(m)=0$

$$\text{得 } L(y) = \frac{L''(m)}{2!}(y - m)^2 + \dots$$

损失函数  $L(y)$  以  $k$  作比例系数, 可近似得  $L(y) \approx k(y - m)^2$

用误差方差  $\sigma^2$  (均方差) 表示测量方法的误差大小。

进行一次测量时,  $L(y) = k\sigma^2$ 。

校正方法有:

- 1) 不校正;
- 2) 零点校正、定点校正;
- 3) 斜率校正(比例校正);
- 4) 一次式校正;
- 5) 曲线校正。

### 2 比例校正与 SN 比

SN 比(Signal - Noise Ratio)就是信号(signal)功率除以噪声(noise)功率的比值。用  $\eta$  表示。

$$\eta = \frac{P_S}{P_N}$$

$$\beta^2 \text{ 的估计值} = \frac{1}{M_1^2 + M_2^2 + \dots + M_n^2} (S_\beta - V_e)$$

测量值的误差方差(估计值)

$$= \frac{V_e}{\frac{1}{M_1^2 + M_2^2 + \dots + M_n^2} (S_\beta - V_e)}$$

测量值误差方差的倒数即为 SN 比  $\eta$ 。

$$\text{测量值的 SN 比 } \eta = \frac{\beta^2}{\sigma^2}$$

$$\text{令 } r = (M_1^2 + M_2^2 + \dots + M_n^2)$$

$$\eta = \frac{\frac{1}{r} (S_\beta - V_e)}{V_e}$$

$$S_\beta = \frac{(M_1 y_1 + M_2 y_2 + \dots + M_n y_n)}{M_1^2 + M_2^2 + \dots + M_n^2} \quad (f=1)$$

$$S_\gamma = y_1^2 + y_2^2 + \dots + y_n^2 \quad (f=k)$$

$$S_e = S_\gamma - T_\beta \quad (f=k-1)$$

$$V_e = \frac{S_e}{k-1}$$

### 3 测量误差及其改善方法

设一个标准的真值为  $M_0$ , 标示值为  $M'$ , 标示误差  $e'$ , 则

$$M' = M_0 + e'$$

并用规定的测量方法读出该标准的读数值为  $y_0$ 。令  $y_0, y$  的回归误差为  $e_0, e$ 。得出观测方程:

$$y_0 = \beta M_0 + e_0$$

$$y = \beta M + e$$

可得比例式校正

$$\sigma_e^2 = \left[ 1 + \left( \frac{y}{y_0} \right)^2 \right] \frac{1}{\eta} + \left( \frac{y}{y_0} \right)^2 \sigma_0^2$$

第一项  $\frac{1}{\eta}$  = 测量方法的误差方差

第二项  $\left( \frac{y}{y_0} \right)^2 \frac{1}{\eta}$  = 伴随校正工作的误差方差

第三项  $\left( \frac{y}{y_0} \right)^2 \sigma_0^2$  = 标准的示值误差给测量值造

成的误差方差

使用比例式校正方法最好选择接近于测量范围最大值的标准。

定点校正

$$\sigma_e^2 = (y - y_0)^2 \sigma_0^2 + 2 \times \frac{1}{2}$$

定点校正减少误差方差的方法:

- 1) 采用接近于平均值的标准  $M_0$ ;
- 2) 减小刻度间隔的误差大小  $\sigma_0^2$ ;
- 3) 减小标准的误差  $\sigma_0^2$ ;
- 4) 加大 SN 比。

减少误差的方法:

- 1) 寻求稳定性好的标准;
- 2) 尽可能准确地确定标准具有的值得;
- 3) 在较好的条件下保存标准;
- 4) 研究 SN 比较大的测量方法;
- 5) 使校正式合理化;
- 6) 使校正周期合理化;

- 7) 进行误差因子校正;
- 8) 测量环境标准化。

#### 4 校正频率与误差变动

误差总是在受到环境的各种变量影响的情况下产生的。假设影响读数值  $y$  的信号因子为  $M$ , 其他变量为  $x_1, x_2, \dots, x_n$ , 误差可表示如下:

$$y = f(M, x_1, x_2, \dots, x_n)$$

进行一次式校正, 以  $M$  的一次式近似表示上式的函数, 即回归误差  $\epsilon$  为

$$\epsilon = f(M, x_1, x_2, \dots, x_n) - (\alpha + \beta M)$$

对各种  $x_1, x_2, \dots, x_n$  的变化。该式平方的平均值就是回归方差分析。

进行测量误差比较时, 应确定信号因子的水平, 这些水平的关键是要把欲比较的测量方法的整个范围都包括进去。

SN 比的比较

以钟表设计方法比较为例, 用  $A_1, A_2$  两种方法设计钟表, 设连续五天  $D_1, D_2, D_3, D_4, D_5$  每三小时进行校正数据如表 46.4-1 所示。

表 46.4-1 钟表的误差数据

时间		A <sub>1</sub>								
		0	3	6	9	12	15	18	21	计
D <sub>1</sub>		0	-1	2	4	5	3	2	-1	14
D <sub>2</sub>		0	-2	5	7	8	4	-2	-1	19
D <sub>3</sub>		1	5	2	4	3	7	3	2	27
D <sub>4</sub>		0	-2	5	3	4	6	-2	-4	10
D <sub>5</sub>		-2	4	3	5	4	2	-1	5	20
计		-1	4	17	23	24	22	0	1	90

时间		A <sub>2</sub>								
		0	3	6	9	12	15	18	21	计
D <sub>1</sub>		-7	-5	-2	0	4	4	7	10	11
D <sub>2</sub>		12	14	17	19	21	24	26	29	162
D <sub>3</sub>		31	33	35	37	40	41	45	47	309
D <sub>4</sub>		49	51	54	57	59	61	63	66	460
D <sub>5</sub>		69	72	74	76	78	86	86	86	627
计		154	165	178	189	202	216	227	238	1569

没有进行任何校正时的误差变动

$A_1$  总误差平方

$$S_T = 0^2 + (-1)^2 + 2^2 + \dots + 5^2 = 550 \quad (f=40)$$

所以误差方差  $V_e$  为

$$V_e = \frac{550}{40} = 13.8$$

对于  $A_2$ ,

$$S_T = 92105 \quad (f=40)$$

$$V_e = \frac{92105}{40} = 2302.6$$

$$A_1, A_2 \text{ 误差大小比为 } \frac{1}{\sqrt{\frac{2302.6}{13.8}}} = 12.9$$

用去日差的方法进行校正时, 误差变动可由总平方和  $S_T$  减去日变动求出。

$A_1$  的日变动

$$S_D = \frac{14^2 + 19^2 + 27^2 + 10^2 + 20^2}{8} = 223.2 \quad (f=5)$$

$$S_e = 550 - 223.2 = 326.8 \quad (f=35)$$

$$V_e = \frac{326.8}{35} = 9.3$$

$A_2$  的日变动

$$S_D = \frac{11^2 + 162^2 + 309^2 + 460^2 + 627^2}{8} = 90821.9$$

( $f=5$ )

$$S_e = 92105 - 90821.9 - 1283.1 \quad (f=35)$$

$$V_e = \frac{1283.1}{35} = 36.7$$

$A_1, A_2$  误差大小比为 1/3.9。

进行一次式校正时, 对于  $A_1$  设  $M_1, M_2, \dots, M_{40}$  为间隔  $3600 \times 3 = 10800$  s 的 40 个水平。对应于各水平有 40 个读数值  $y_1, y_2, \dots, y_{40}$ 。

$$S_m = \frac{(\text{合计})^2}{40} = 202.5 \quad (f=1)$$

回归变动  $S_\beta$  可以由信号因子  $M$  与读数值  $y$  的共同变动  $S(My)$  的平方除以信号因子的水平间变动  $S(MM)$  求得。

$$\begin{aligned} S(MM) &= M_1^2 + M_2^2 + \dots + M_{40}^2 - \frac{(M_1 + M_2 + \dots + M_{40})^2}{40} \\ &= 0^2 + 3^2 + \dots + 117^2 - \\ &\quad \frac{(0 + 3 + \dots + 117)^2}{40} = 47970 \end{aligned}$$

$$\begin{aligned} S(My) &= M_1 y_1 + M_2 y_2 + \dots + M_{40} y_{40} - \\ &\quad \frac{(M_1 + M_2 + \dots + M_{40})(y_1 + y_2 + \dots + y_{40})}{40} \\ &= 87 \end{aligned}$$

所以

$$S_\beta = \frac{S(My)^2}{S(MM)} = 0.2$$

求出误差变动和误差方差

$$\begin{aligned} S_e &= S_T - S_m - S_\beta = 550 - 202.5 - 0.2 \\ &= 347.3 \quad (f=38) \end{aligned}$$

$$V_e = \frac{347.3}{38} = 9.1$$

$S_\beta$  最好根据表 4.1 中对应 0, 3, 6, ..., 21 的合计栏来求, 将水平  $M$  每天的 0, 3, 6, ..., 21 改为  $M_1, M_2, M_3, \dots, M_8$ , 对应于  $M_1, M_2, M_3, \dots, M_8$  各有五个数据之和, 所以  $M_8$  可以由下式求:

$$S_\beta = \frac{[(M_1 - \bar{M})Y_1 + (M_2 - \bar{M})Y_2 + \dots + (M_8 - \bar{M})Y_8]^2}{5[(M_1 - \bar{M})^2 + (M_2 - \bar{M})^2 + \dots + (M_8 - \bar{M})^2]}$$

式中  $Y_1, Y_2, \dots, Y_8$  是分别对应于  $M_1, M_2, \dots, M_8$

的每五个数据之和。

对于 A, 可求得  $S_f=0.1$ , 此时误差变动为:

$$S_r = S_f - S_D - S_a = 326.7 \quad (f=34)$$

$$V_r = \frac{326.7}{31} = 9.6$$

对于 A, 可得

$$S_r = 1250.7 \quad (f=1)$$

$$S_e = S_r - S_D - S_a = 32.4 \quad (f=34)$$

$$V_r = 0.95$$

一次式校正公式

$$\eta = \frac{\frac{1}{r}(S_s - S_r)}{V_r}$$

当信号因子的水平也用与读数值同样的因次与单位时, 如果  $\beta \approx 1$  上式的分子约为 1, 所以 SN 比可表示为

$$\eta \approx \frac{1}{V_r}$$

进行 SN 比的比较时的关键:

1) 对重要的误差因子, 要预测它在校正周期内的变化程度, 定出水平;

2) 对于定点校正或者一次式校正, 从总平方和中应减掉的变动不同。

## 5 信号因素水平的确定方法

### 5.1 化学分析的例子

化学分析研究的重点就是研究稀浓物质的定量法。在化学分析的情况下可以将欲测的分子或原子浓度大致清楚的物质用各种比例进行混合, 来确定信号因子的水平。

例如为测定酱油中的总氮量, 比较两种方法, 控制因子 A 的水平, A<sub>1</sub>: 基耶达法, A<sub>2</sub>: 新方法, 确定信号因子的问题。(NH<sub>4</sub>)<sub>2</sub>SO 的溶液制成 0mgN/mL、5mgN/mL、10mgN/mL、15mgN/mL、20mgN/mL, 加进欲测的 90mL 酱油, 制成五水平的信号因子。假设欲测酱油的含氮量真值为 a' (未知), 当 a = 0.9a' 时, 五个信号因子的水平 M 如下:

$$M_1 = a(\%)$$

$$M_2 = a - h(\%)$$

$$M_3 = a + 2h(\%)$$

$$M_4 = a + 3h(\%)$$

$$M_5 = a + 4h(\%)$$

间隔 h 是未知的, 由两名分析人员分别对五水平的信号因子 M<sub>1</sub>、M<sub>2</sub>、M<sub>3</sub>、M<sub>4</sub>、M<sub>5</sub> 进行测定, 得数据, 如表 46.4-2 所示。

表 46.4-2 两名分析人员 R<sub>1</sub>、R<sub>2</sub>

进行总氮量分析的读数值

方法 A<sub>1</sub>

	M <sub>1</sub>	M <sub>2</sub>	M <sub>3</sub>	M <sub>4</sub>	M <sub>5</sub>	计
R <sub>1</sub>	22.12	24.71	29.37	32.98	36.42	
R <sub>2</sub>	22.15	25.57	29.05	32.77	36.31	
计	44.27	50.28	58.42	65.75	72.73	291.45

方法 A<sub>2</sub>

	M <sub>1</sub>	M <sub>2</sub>	M <sub>3</sub>	M <sub>4</sub>	M <sub>5</sub>	计
R <sub>1</sub>	18.02	21.29	24.44	27.69	30.94	
R <sub>2</sub>	18.04	21.20	24.30	27.38	30.92	
计	36.06	42.49	48.74	55.07	61.86	244.22

方差分析:

下面两个回归式

$$y = a + \beta M = a - \beta \bar{M} + \beta(M - \bar{M})$$

$$y = \beta M = \beta \bar{M} + \beta(M - \bar{M})$$

用于方差分析的回归项  $\beta(M - \bar{M})$  与残差项的变动。

可以计算总变动 S<sub>T</sub>

$$S_T = \text{数据的总平方和} \left( - \frac{\text{数据和的平方}}{\text{数据个数}} \right)$$

(f = 数据个数)

回归项的变动为

$$S_p = \frac{(M_1 y_1 + M_2 y_2 + \dots + M_k y_k)}{r}$$

$$r = r_1 M_1^2 + r_2 M_2^2 + \dots + r_k M_k^2$$

r<sub>1</sub>, r<sub>2</sub>, ..., r<sub>k</sub> 每个水平的重复数

y<sub>1</sub>, y<sub>2</sub>, ..., y<sub>k</sub> 信号因子各水平的合计

$$\eta = \frac{\frac{1}{r}(S_p - V_r)}{V_e}$$

误差变动 S<sub>e</sub> = S<sub>T</sub> - S<sub>p</sub>

误差的标准方差 s 是

$$S = \frac{1}{\sqrt{\eta}}$$

代入数据, 计算过程略, 可得 A<sub>1</sub>, A<sub>2</sub> 的方差分析表(如表 46.4-3、表 46.4-4 所示)

表 46.4-3 A<sub>1</sub> 的方差分析表

来源	f	S	V
β	1	262.0156	262.0156
e	8	0.8852	0.1106
T	9	262.9008	

表 46.4-4 A<sub>2</sub> 的方差分析表

来源	f	S	V
β	1	205.9536	205.9536
e	8	0.1058	0.0132
T	9	206.0594	

5.2 质量测定的例子

在一般物理测定中是不能像化学测试那样混合，但可以相加，例质量为  $\alpha$  的物质与质量为  $\beta$  的物质相加后得质量为  $(\alpha + \beta)$  的物质。

现对婴儿专用秤  $A_1$  与普通秤  $A_1$  进行 SN 比的实验，婴儿专用秤  $A_1$  的计量范围为 0~20kg。为求整个计量范围的误差方差，用以下四种物品：

- A: 约 10kg 的物品
- B: 约 5kg 的物品
- C: 约 3kg 的物品
- D: 约 1kg 的物品

设上述物品未放上记为 1，放上为 2，对应 16 种组合秤的读数值为  $y_1, y_2, \dots, y_{16}$ ，如表所列

求 A、B、C、D 四种物品的变动  $S_A, S_B, S_C, S_D$ ，然后由总变动减去这些变动从而得到误差变动。所秤某一物品的变动可由没放上时的 8 个数据之和与放上物品时 8 个数据之和的差值平方后，除以 16 求得。

求 SN 比：

$$\eta = \frac{1}{r} \frac{(S_A + S_B - S_C + S_D - 4V_e)}{V_e}$$

$r = A, B, C, D$  的真值变动，可通过估计值  $M_A, M_B, M_C, M_D$  求得。

$$r = 8(\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_D^2) = 4(M_A^2 + M_B^2 + M_C^2 + M_D^2)$$

计算得方差分析表(见表 46.4-5 和表 46.4-6)。

真值不明时的信号因子水平的确定方法，测量标准品时与测量实际物品相比较测量误差比较大时，应

表 46.4-5  $A_1$  的方差分析表

来源	f	S	V
A	1	374.51926	374.51926
B	1	79.52181	79.52181
C	1	31.50016	31.50016
D	1	3.35806	3.35806
e	11	0.00355	0.000323
T	15	488.90284	

表 46.4-6  $A_2$  的方差分析表

来源	f	S	V
A	1	369.1202	
B	1	73.3164	
C	1	31.7814	
D	1	3.6577	
e	11	0.1129	0.00103
T	15	477.9886	

针对具体问题进行分析研究确定实际测量对象的信号因子水平。

5.3 以轮圈动平衡测量为例子

设不平衡量小、中、大三个轮圈的不平衡量  $M$  分以下三水平：

$$\left. \begin{aligned} M_1 &= M' \\ M_2 &= M' + h \\ M_3 &= M' + 2h \end{aligned} \right\}$$

控制因子为两台平衡装置  $A_1$  和  $A_2$ ，按照各个水平，求出与  $M_1, M_2, M_3$  对应的合计值以及全部合计值，见表 46.4-7。

表 46.4-7 不平衡量的读数值  
平衡装置  $A_1$  的情形

	$M'_1$	$M' + h$	$M'_1 + 2h$	$M'_2$	$M'_2 + h$	$M'_2 + 2h$	$M'_3$	$M'_3 + h$	$M'_3 + 2h$
$R_1$	0.20	0.60	1.10	0.80	1.20	1.80	1.00	1.60	2.10
$R_2$	0.10	0.60	1.20	0.70	1.20	1.80	1.00	1.60	2.20
计	0.30	1.20	2.30	1.50	2.40	3.60	2.00	3.20	4.30
	$M'$			$M' + h$			$M' + 2h$		
计	3.80			6.80			10.20		
合计									
20.80									

平衡装置  $A_2$  的情形

	$M'_1$	$M' + h$	$M'_1 + 2h$	$M'_2$	$M'_2 + h$	$M'_2 + 2h$	$M'_3$	$M'_3 + h$	$M'_3 + 2h$
$R_1$	0.25	1.00	1.60	1.10	1.30	1.90	1.50	1.90	2.70
$R_2$	0.60	0.70	1.40	1.20	1.10	2.00	1.35	1.80	2.20
计	0.85	1.70	3.00	2.30	2.40	3.90	2.85	3.70	4.90
	$M'$			$M' + h$			$M' + 2h$		
计	6.00			7.80			11.80		
合计									
25.60									

详细计算  $A_1$ , 省略  $A_2$  计算。

$$S_T = 0.20^2 + 0.10^2 + 0.60^2 + \dots + 2.20^2 = 30.2800 \quad (f=18)$$

$$S_M = \frac{3.80^2 + 7.50^2 + 9.50^2}{6} = 26.8233 \quad (f=3)$$

$$S_\beta = \frac{(-3.80 + 10.20)^2}{12} = 3.4133 \quad (f=1)$$

$$S_e = S_T - S_M - S_\beta = 0.0434 \quad (f=14)$$

得出方差分析表(见表 46.4-8 和表 46.4-9)。

表 46.4-8  $A_1$  的方差分析表与 SN 比的计算

来源	$f$	$S$	$V$
$M'$	3	26.8233	—
$\beta$	1	3.4133	3.4133
$e$	14	0.0434	0.00310
$T$	18	30.2800	

表 46.4-9  $A_2$  的方差分析表与 SN 比的计算

来源	$f$	$S$	$V$
$M'$	3	39.3108	—
$\beta$	1	2.8033	2.8033
$e$	14	0.5309	0.0379
$T$	18	42.6450	

对于  $A_1$ :

$$\eta = \frac{\frac{1}{r}(S_\beta \cdot V_e)}{V_e} = \frac{1}{6 \times 2 \times 15^2} (3.4133 \cdot 0.00310) = 0.407 = 3.90(\text{dB})$$

$$= 0.407 = 3.90(\text{dB})$$

对于  $A_2$ :

$$\eta = 0.0270 = -15.7(\text{dB})$$

$$\text{上二式取: } 10 \cdot \log 0.407 = -3.90(\text{dB})$$

$$10 \cdot \log 0.0270 = -15.7(\text{dB})$$

在进行全数据测量、修正和调整时,减少测量误差和减少修正或调整后物品偏差,会产生同样的效果。

用回归变量(操作变量)求信号因素的水平。

等间隔的信号因子水平:

设物品长度根据室温变化,室温可控制到  $\pm 0.1^\circ\text{C}$ ,将物品定为  $M_1 = 5^\circ\text{C}$ 、 $M_2 = 15^\circ\text{C}$ 、 $M_3 = 25^\circ\text{C}$ 、 $M_4 = 35^\circ\text{C}$  四水平,用以一天为单位的测量方法求出读数值。现设  $R_1, R_2, R_3$  在三天进行这实验,数据如表 46.4-10 和表 46.4-11 所示,其中物品的长度大约为  $146352\mu\text{m}$ 。

表 46.4-10 读数值的数据

(已减去假定平均值)

	$M_1$	$M_2$	$M_3$	$M_4$
$R_1$	0.053	0.125	0.198	0.270
$R_2$	0.040	0.111	0.183	0.251
$R_3$	0.054	0.123	0.200	0.273

将数据放大 1000 倍,读数值单位  $\mu\text{m}$ 。

数据分析:

表 46.4-11 经过加工的数据

	$M_1$	$M_2$	$M_3$	$M_4$	计
$R_1$	53	125	198	270	646
$R_2$	40	111	183	251	585
$R_3$	54	123	200	273	650
计	147	359	581	794	1881

总变动  $S_T$  为

$$S_T = 53^2 + 125^2 + \dots + 273^2 - \frac{1881^2}{12} = 78676 \quad (f=)$$

11)

信号的变动根据  $k=4$  的等间隔公式求解,得

$$S_\beta = \frac{(3M_1 - M_2 + M_3 + 3M_4)}{r_0 \times 20} = 77976 \quad (f=1)$$

误差变动  $S_e$  为

$$S_e = S_T - S_\beta = 700 \quad (f=10)$$

得误差方差  $V_e = 70.0$

当已知的误差在 10% 以下时,求出

$$\beta = \frac{-3 \times 147 - 359 + 581 + 3 \times 794}{r_0 \lambda S h} = \frac{2163}{3 \times 10 \times 10} = 7.21(\text{nm}/^\circ\text{C})$$

温度间隔  $10^\circ\text{C}$  时的尺寸间隔  $h$  为  $h = h\beta = 10 \times$

$$7.21 = 72.1(\text{nm})$$

求 SN 比:

$$\eta = \frac{\frac{1}{r}(S_\beta - V_e)}{V_e} = \frac{1}{r_0 \times 5 \times h^2} (S_\beta - V_e) = 0.014273 = -18.46(\text{dB})$$

设  $R_1$  为  $R_i$  天的数据之和,求日间的变动  $S_R$

$$S_R = \frac{R_1^2 + R_2^2 + R_3^2}{4} - \frac{T^2}{12} = \frac{646^2 + 585^2 + 650^2}{4} - \frac{1881^2}{12} = 664 \quad (f=2)$$

$R_1$  的一次系数为

$$\frac{-3 \times 53 - 125 + 198 + 3 \times 270}{r_0 \times 10 \times h} = 7.24(\text{nm}/^\circ\text{C})$$

为了变动的简便计算,省略分母之后的一次系数(放大了 100 倍),

$$L(R_1) = 724$$

同样,

$$L(R_2) = 705$$

$$L(R_3) = 734$$

三个估计值  $L(R_1), L(R_2), L(R_3)$  之间的变动为:

$$S_{R \times \beta} = \frac{L(R_1)^2}{\text{单位数}} + \frac{L(R_2)^2}{\text{单位数}} + \frac{L(R_3)^2}{\text{单位数}} - \frac{[L(R_1) + L(R_2) + L(R_3)]^2}{\text{单位数}}$$

分母的单位数为平方之前估计值的单位数。

单位数的求解方法:

设  $n$  个观测值的  $y_1, y_2, \dots, y_n$  的常系数线性式为  $L$ , 则

$L = c_1 y_1 + c_2 y_2 + \dots + c_n y_n$  的单位数即为系数的平方和  $D$ , 即

$$D = c_1^2 + c_2^2 + \dots + c_n^2$$

$$S_{R \times \beta} = \frac{724^2 + 705^2 + 734^2}{20} - \frac{(724 + 705 + 734)^2}{60} =$$

22 ( $f=2$ )

误差变动  $S_e$  为

$$S_e = S_T - S_R - S_\beta - S_{R \times \beta} = 78676 - 77976 - 664 - 22 = 14$$

作得方差分析表(见表 46.4-12)。

表 46.4-12 方差分析表

来源	自由度 $f$	变动 $S$	方差 $V$	方差的期望值 $E(V)$	纯变动 $S'$	有效率 $\rho(\%)$
$\beta$	1	77976	77976	$\sigma^2 + 5r_0 h^2 \beta^2$	77973.7	99.107
$R$	2	664	332	$\sigma^2 + 4\sigma_R^2$	659.4	0.838
$\beta \times R$	2	22	11	$\sigma^2 + 5h^2 \sigma_{\beta \times R}^2$	17.4	0.022
$e$	6	14	2.3	$\sigma^2$	25.5	0.033
$T$	11	78676			78676.0	100.000

回归量不好控制的情况

表 46.4-13 所示温度未能很好控制,  $x$  为温度,  $y$  为长度的读数值。

表 46.4-13 温度  $x(^{\circ}\text{C})$  与读数值  $y$

		(nm)				
		$M_1'$	$M_2'$	$M_3'$	$M_4'$	计
$R_1$	$x$	5.8	15.1	25.0	37.6	83.5
	$y$	61	128	199	291	679
$R_2$	$x$	4.3	17.9	25.9	33.8	81.9
	$y$	43	140	198	256	637
$R_3$	$x$	6.3	13.8	27.8	37.6	85.5
	$y$	68	123	223	294	708

由  $R_1, R_2, R_3$  分别求一次系数, 由线性式  $L =$

$$\sum_{i=1}^4 (x_i - \bar{x}) y_i, \text{ 可知 } D = \sum_{i=1}^4 (x_i - \bar{x})^2$$

$$(R_1) L_1 = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 -$$

$$\frac{\sum x_i \sum y_i}{4} = 4029.075$$

$$D_1 = x_1^2 + x_2^2 + x_3^2 + x_4^2 - \frac{(\sum x)^2}{4} = 557.3475$$

$$(R_2) L_2 = 3429.325$$

$$D_2 = 475.2475$$

$$(R_3) L_3 = 4246.100$$

$$D_3 = 589.1675$$

$$S_\beta = \frac{(4029.075 + 3429.325 + 4246.100)^2}{557.3475 + 475.2475 + 589.1675} = 84473.11$$

$$S_T = 61^2 + 128^2 + \dots + 294^2 - \frac{2024^2}{12} = 85112.67$$

( $f=11$ )

$$S_R = \frac{679^2 + 637^2 + 708^2}{4} - \frac{2024^2}{12} = 637.17 \quad (f=2)$$

$$S_{\beta \times R} = \frac{L_1^2}{D_1} + \frac{L_2^2}{D_2} + \frac{L_3^2}{D_3} - \frac{(L_1 + L_2 + L_3)^2}{D_1 + D_2 + D_3}$$

$$= \frac{4029.075^2}{557.3475} + \frac{3429.325^2}{475.2475} + \frac{4246.100^2}{589.1675} -$$

$$\frac{11704.5^2}{1621.7625}$$

$$= 0.14 \quad (f=2)$$

作出方差分析表(见表 46.4-14)。

表 46.4-14 方差分析表

来源	$f$	$S$	$V$
$\beta$	1	84473.11	84473.11
$R$	2	637.17	318.58
$\beta \times R$	2	0.14	0.07
$e$	6	2.25	0.38
$T$	11	85112.67	

SN 比计算只需  $\beta$  和  $e$  的方差期望值,  $\beta^2$  需要的系数是单位数  $D, D = D_1 + D_2 + D_3$ 。

另外小于误差方差 1/2 的项一般归入误差中, 而不再求纯变动。作出经整理的方差分析表(如表 46.4-15 所示)。

表 46.4-15 经过整理的方差分析表

来源	$f$	$S$	$V$	$E(V)$	$S'$	$\rho(\%)$
$\beta$	1	84473.11	84473.11	$\sigma^2 + D\beta^2$	84472.81	99.248
$R$	2	637.17	318.58	—	636.57	0.748
$e$	8	2.39	0.30	$\sigma^2$	3.29	0.004
$T$	11	85112.67			85112.67	100.000

SN 比中的  $r$  由

$$r = D(\hat{\beta})^2 = S_\beta$$

$$\hat{\beta} = \frac{11704.5}{1621.7625} = 7.2171$$

$$\eta = \frac{1}{84473.11} (84473.11 - 0.30)$$

$$= 3.33 = 5.22(\text{dB})$$

用简易法求得误差范围

$$\pm \frac{3}{\sqrt{3.33}} = \pm 1.6(\text{nm})$$

## 6 误差因素的选取方法

误差因素指在读数值时所受到的除信号因子以外的变量, 分因环境条件变量影响读数值的外部随机误差与因测量系统因素的变化而使读数值变化的内部随机误差两种。

在校正周期内产生的内部随机误差和外部随机误差的变化是无法用校正消除的。所以, 应注意以下几点:

1) 如果标准稳定性不好, 那么一般都得不到好的

测量结果。在以实物样品为标准时要充分注意。

2) 在标准的真值或内部随机误差的塑性变化较小的情况下,如果用标准条件进行测试,则也可以认为是稳定性好的标准。但是当物品在现场和标准室中发生弹性变形时,其真值不是现场测量的值而是在标准室测得的值。

3) 即使用稳定性好的标准或者几乎不存在塑性变化的标准,在标准条件下测量时,还要评价测量系统误差的问题。即使知道标准条件下的误差大小仍不知道现场测试时的误差大小。

4) 测量系统的内部随机误差和外部随机误差虽可以用标准进行校正,但是校正周期内的变化则作为 SN 比的误差。

5) 物品真值的时间、空间变化是真值的定义问题。

6) 标准的真值变化或内部随机变化作为标准的误差必须与 SN 比误差分开研究。

## 7 改善 SN 比的实验设计

### 7.1 信号因素与误差因素

信号因子往往只有一个而误差因子常常多个同时存在。假设信号因子  $M$  以及误差因子  $K, L, N, Q, R$  都取三水平。三水平系的正交表适合于进行测量方法的研究。常用的有  $L_9, L_{18}, L_{27}, L_{33}$  等正交表。

$L_{18}$  的第一列为八个水平,因子的安排

1) 当信号因子  $M$  为三水平时,第一水平对应于表  $L_{18}$  第一列的 1, 2, 第二水平对应于第一列的 3, 4, 第三水平对应于 5, 6。

2) 当  $M$  为四水平时,将  $M_1$  和  $M_3$  重复安排。即  $M_1$  对应于 1, 2

$M_2$  对应于 3

$M_3$  对应于 4

$M_4$  对应于 5, 6

为拟水平法。

3) 当为五水平时,给  $M_1, M_2, M_3, M_4, M_5$  其中一个(通常  $M_5$ )加入虚拟水平。

4)  $M$  为六水平时,按表进行布置。

其余 2~7 列安排误差因子。

### 7.2 SN 比的计算

方差分析比较简单,只需求出总变动和信号因素的效应。

通过求得误差变动  $S_e$  与误差方差  $V_e$  后,求 SN 比:

$$\eta = \frac{\frac{1}{r}(S_f - V_e)}{V_e}$$

SN 比本身是误差方差的倒数,要减小误差方差,就要提高信号因素的灵敏度,或者相对减弱误差因素的效应。

### 7.3 误差因素及其控制措施

通过方差分析即可以搞清哪个误差因子对读数值的影响大。

SN 比的实验设计方法,就是希望根据 SN 比的数值求出直接的误差方差,就是多选取改善 SN 比的因子即控制因子,然后将其安排在正交表中。

#### 7.3.1 内部正交表——控制因素的安排

实验设计步骤:

1) 选取控制因素,进行多元配置或者安排到一张正交表中。

2) 将信号因子与误差因子一起采用与上述不同的多元配置或者一张不同的正交表进行安排。

3) 求出 1) 和 2) 所有组合的读数值。

4) 按照 1) 中所列的各个实验序号求与 2) 中安排对应的 SN 比,用分贝表示。

5) 按 1) 所排的各个实验序号进行数据分析,即把求出 SN 比作为目标特性的数据。以后步骤与一般实验相同。

安排控制因素的正交表大多采用二水平的正交表,常用  $L_8, L_{16}, L_{32}$  三种也可以采用三水平的正交表。

因为对于安排了控制因素的正交表的各个实验序号想把 SN 比的值作为目标特性来求,所以必须按照求 SN 比的正交表进行安排,将此表称为内部正交表。要安排交互作用需要使用点线图。

#### 7.3.2 外部正交表——信号因素的安排

安排了控制因素的正交表为内部正交表,而安排了信号因子和误差因子的正交表为外部正交表,外部正交表一般采用三水平的正交表。在内部正交表  $L_8$  中安排控制因素,在外部正交表  $L_9$  中安排信号因子与误差因子,再对两者的组合条件进行实验,称之为正交表的  $L_8$  和  $L_9$  的直积实验。

对于测量方法的改进实验,当内部正交实验为  $L_8$ , 外部正交实验  $L_9$  时,需要对两者的全部组合  $MN$  个条件求出读数值。

## 8 求解标准固有误差的实验设计方法

### 8.1 标准的误差

在进行测量研究时首先是定义国际的绝对标准。

其次在此基础上规定各个阶段的标准,然后应用这些标准进行测量方法研究。

标准的稳定性就是标准保持真值的时间变化性。就是要求出标准的时间变化大小,即用已求出的SN比的测量方法按不同时间测量标准。设测量值为  $y_1, y_2, \dots, y_n$ , 则  $y$  之间的变动  $S_T$  为

$$S_T = y_1^2 + y_2^2 + \dots + y_n^2 - \frac{(y_1 + y_2 + \dots + y_n)^2}{n}$$

$V_T$  的期望值是标准的变动  $\sigma M^2$  与测量方法的方差  $\sigma^2$  之和。即

$$E(V_T) = \sigma M^2 + \sigma^2$$

标准研究中需要选取误差因子如温度、湿度等环境条件及随时间变化而产生的劣化,设  $K, L, \dots$  等。

### 8.2 直读天平的实验

实验是在不知道直读天平误差的条件进行的,所以是求解直读天平误差SN比误差问题和求解砝码的误差问题同时进行,为此按照求解标准的误差方差安排和求解SN比误差的安排方法进行二者的直积实验。实验目的就是要求出200g量程的天平本身的标准即砝码的误差的平均值。取作信号因子的  $A, B, C, D, E$  五个物品质量大致为:  $A = 6.2g, B = 14.5g, C = 23.4g, D = 48.8g, E = 92.9g$ 。

选取正交表  $L_{32}$ , 将这些组合安排到第1、2、4、8、16列,因而得出了具有32水平的信号因子。准备了  $M', M'', M'''$  三个微调砝码,标示值分别为10mg, 5mg, 2mg。设放上为第一水平,放上时为第二水平安排如表46.4-16所示。

表 46.4-16 信号因子的安排

No.	$M'$	$M''$	$M'''$
1	1	1	1
2	1	1	2
3	1	2	1
4	1	2	2
5	2	1	1
6	2	1	2
7	2	2	1
8	2	2	2

因子  $R$  表示  $R_1, R_2, R_3, R_4$  四天里进行这一实验,各天的误差  $R$  安排到正交表  $L_{32}$  第3、5、6列作出的四水平中,由3、5、6列的四水平的组合与四水平的  $R$  对应如下:

- 1 1 1 → 1
- 1 2 2 → 2
- 2 1 2 → 3

2 2 1 → 4

把  $R_1$  分 No. 1~No. 4, ..., No. 29~No. 32 八组实验,对每组进行校正就是要求出SN比的误差。组内的零点偏移归入误差中,所列数据是质量的读数值减去  $A, B, C, D, E$  的质量估计值后的数据。

### 8.3 方差分析

先求总变动。

$$CF = \frac{11462.6^2}{256} = 513246.870 \quad (f=1)$$

$$S_T = 0.0^2 + 2.1^2 + 5.1^2 + \dots + 89.0^2 - CF = 5547.090 \quad (f=255)$$

对  $M', M'', M'''$  的八种组合,求出32个数据的总和,然后  $M', M'', M'''$  各水平的和,将第一、二水平之差的平方除以256,即三个微调砝码效应的大小。

$$S_M = \frac{(5089.4 - 6373.2)^2}{256} = 6438.056 \quad (f=1)$$

$$S_{M'} = \frac{(5410.0 - 6052.6)^2}{256} = 1613.026 \quad (f=1)$$

$$S_{M''} = \frac{(5602.4 - 5860.2)^2}{256} = 259.613 \quad (f=1)$$

$$S_{T_1} = \frac{1}{8} (68.7^2 + 206.6^2 + \dots + 644.4^2) - CF = 67235.550 \quad (f=31)$$

$$S_A = \frac{(4983.1 - 6479.5)^2}{256} = 8746.926 \quad (f=1)$$

$$S_B = \frac{(4884.9 - 6577.7)^2}{256} = 11193.640 \quad (f=1)$$

$$S_C = \frac{(4717.4 - 6745.2)^2}{256} = 16062.394 \quad (f=1)$$

$$S_D = \frac{(4839.8 - 6622.8)^2}{256} = 12418.316 \quad (f=1)$$

$$S_E = \frac{(4634.7 - 6827.9)^2}{256} = 18789.556 \quad (f=1)$$

$$S_R = \frac{2855.6^2 + 2874.3^2 + 2853.1^2 + 2879.6^2}{64} - CF = 8.249 \quad (f=3)$$

砝码间的误差

$$S_{e_1} = S_T - (S_A + S_B + S_C + S_D + S_E + S_R) = 16.641 \quad (f=32)$$

二次误差变动

$$S_{e_2} = S_T - (S_{T_1} + S_M + S_{M'} + S_{M''}) = 0.845 \quad (f=221)$$

作出方差分析表46.4-17。



表 46.4-17 方差分析表

来源	<i>f</i>	<i>S</i>	<i>V</i>	<i>E(V)</i>
<i>T</i> <sub>1</sub>	<i>A</i>	8746.926	6746.926	$\sigma_0^2 - 3\sigma_1^2 + 128\sigma_A^2$
	<i>B</i>	11193.640	11193.640	$\sigma_0^2 + 8\sigma_1^2 - 128\sigma_B^2$
	<i>C</i>	16062.394	16062.394	$\sigma_0^2 - 8\sigma_1^2 + 128\sigma_C^2$
	<i>D</i>	12418.316	12418.316	$\sigma_0^2 + 8\sigma_1^2 + 128\sigma_D^2$
	<i>E</i>	18789.556	18789.556	$\sigma_0^2 - 8\sigma_1^2 + 128\sigma_E^2$
<i>R</i>	3	8.249	2.750	$\sigma_0^2 + 8\sigma_1^2 + 240\sigma_R^2$
<i>e</i> <sub>1</sub>	23	16.641	0.724	$\sigma_0^2 + 8\sigma_1^2$
<i>M'</i>	1	6438.056	6438.056	$\sigma_0^2 + 128\sigma_{M'}^2$
<i>M''</i>	1	1613.026	1613.026	$\sigma_0^2 + 128\sigma_{M''}^2$
<i>M'''</i>	1	259.613	259.613	$\sigma_0^2 + 128\sigma_{M'''}^2$
<i>e</i> <sub>2</sub>	221	0.845	0.0038	$\sigma_0^2$
<i>T</i>	255	76647.090		

8.4 标示值误差的问题与数据分析

设 *k* 个标准为 *M*<sub>1</sub>, *M*<sub>2</sub>, ..., *M*<sub>*k*</sub>, 其表示值为 *M*'<sub>1</sub>, *M*'<sub>2</sub>, ..., *M*'<sub>*k*</sub>. 对 *M*<sub>1</sub>, *M*<sub>2</sub>, ..., *M*<sub>*k*</sub> 随机用某种测量方法各测一次。如此反复进行 *r* 次。设反复数为 *R*<sub>1</sub>, *R*<sub>2</sub>, ..., *R*<sub>*k*</sub>. 得到数据如表 46.4-18 所示。

表 46.4-18 读数值的数据

	<i>M</i> <sub>1</sub>	<i>M</i> <sub>2</sub>	...	<i>M</i> <sub><i>k</i></sub>	计
<i>R</i> <sub>1</sub>	<i>y</i> <sub>11</sub>	<i>y</i> <sub>12</sub>	...	<i>y</i> <sub>1<i>k</i></sub>	<i>R</i> <sub>1</sub>
<i>R</i> <sub>2</sub>	<i>y</i> <sub>21</sub>	<i>y</i> <sub>22</sub>	...	<i>y</i> <sub>2<i>k</i></sub>	<i>R</i> <sub>2</sub>
⋮	⋮	⋮	⋮	⋮	⋮
<i>R</i> <sub><i>r</i></sub>	<i>y</i> <sub><i>r</i>1</sub>	<i>y</i> <sub><i>r</i>2</sub>	...	<i>y</i> <sub><i>r</i><i>k</i></sub>	<i>R</i> <sub><i>k</i></sub>

求出表中横向纵向的总和, 然后求出总变动 *S*<sub>*T*</sub>, *M* 间的变动 *S*<sub>*M*</sub> 和 *R* 间的变动 *S*<sub>*R*</sub>.

$$S_T = y_{11}^2 + y_{12}^2 + \dots + y_{rk}^2 - \frac{T^2}{rk} \quad (f = kr - 1)$$

$$S_M = \frac{M_1^2 + M_2^2 + \dots + M_k^2}{r} - \frac{T^2}{rk} \quad (f = k - 1)$$

$$S_R = \frac{R_1^2 + R_2^2 + \dots + R_k^2}{k} - \frac{T^2}{rk} \quad (f = r - 1)$$

$$S_e = S_T - (S_M + S_R) \quad (f = (k - 1)(r - 1))$$

$$S_\beta = \frac{[\sum (M'_i - M') M_i]^2}{r \sum (M'_i - M')^2} \quad (f = 1)$$

$$S_e = S_M - S_\beta \quad (f = k - 2)$$

方差分析表如表 46.4-19 所示。

表 46.4-19 方差分析表

来源	<i>f</i>	<i>S</i>	<i>V</i>	<i>E(V)</i>
<i>M</i>	$\beta$	1	<i>S</i> <sub><math>\beta</math></sub>	<i>V</i> <sub><math>\beta</math></sub>
	<i>e'</i>	<i>k</i> -2	<i>S</i> <sub><i>e'</i></sub>	<i>V</i> <sub><i>e'</i></sub>
<i>R</i>	<i>r</i> -1	<i>S</i> <sub><i>R</i></sub>	<i>V</i> <sub><i>R</i></sub>	$\sigma^2 + k\sigma_0^2$
<i>e</i>	( <i>k</i> -1)( <i>r</i> -1)	<i>S</i> <sub><i>e</i></sub>	<i>V</i> <sub><i>e</i></sub>	$\sigma^2$
<i>T</i>	<i>kr</i> -1	<i>S</i> <sub><i>T</i></sub>		

在刻度间隔基本正确的前提下, 标准标示值的误差方法可用下式表示:

$$\hat{\sigma}_0^2 = \frac{1}{r} (V_e - V_\beta)$$

9 标准稳定性的实验设计与数据分析

9.1 标准稳定性

稳定性好就是这个标准的真值不随时间和环境的变化而变化。

信号因子指使目标特性即标准的真值趋近于给定值的因子。用符号 *M* 表示标准真值的数值, 用 *Q* 表示修正真值的信号因子, 把 *Q* 称为调整因子。

9.2 稳定性的实验设计

没有调整因子的情况是指仅有误差因子作为环境条件的因子。设温度、湿度和尘埃分别为 *D*、*E*、*F*, 且都为三水平, 温度与湿度有关, 对温度取下列三水平:

$$D_1 = \frac{1}{2} (\text{最低温度} + \text{平均温度})$$

$$D_2 = \text{平均温度}$$

$$D_3 = (\text{最高温度} + \text{平均温度})$$

设温度平均值为 *m*, 标准偏差为  $\sigma^2$ , 则三水平改写为:

$$D_1 = m - \sqrt{\frac{3}{2}} \sigma$$

$$D_2 = m$$

$$D_3 = m + \sqrt{\frac{3}{2}} \sigma$$

温度接近于 *D*<sub>2</sub> 时的湿度分布:

$$E_1 = \frac{1}{2} (\text{最低湿度} + \text{平均湿度})$$

$$E_2 = \text{平均湿度}$$

$$E_3 = (\text{最高湿度} + \text{平均湿度})$$

误差因子 *F* 也同様。

劣化问题需取时间因素, 假设设计寿命保证五年时间内不劣化, 更希望五年后的误差仍比较小。劣化问题于使用条件与使用频率有关。一般取使用频率为误差因子 *G*, 取如下水平。

*G*<sub>1</sub>: 较少频率

*G*<sub>2</sub>: 一般频率

*G*<sub>3</sub>: 较多频率

因为把劣化作为问题提出, 所以规定至少要取设计寿命 5 年的 1/10 即 6 个月作为时间因素 *K* 的水平取择:

*K*<sub>1</sub>: 实验开始

*K*<sub>2</sub>: 3 个月后

*K*<sub>3</sub>: 6 个月后

可以把 *K* 的效应分解为一次效应 *K*<sub>1</sub> 和二次效应 *K*<sub>2</sub>。将误差因子 *D*、*E*、*F*、*G* 安排在正交表 *L*<sub>9</sub> 中, 用九种

条件进行实验。标准的种类有  $A_1, A_2$  两种, 进行稳定性比较。用  $A_1, A_2$  方法各制出九个给定值为  $M_i$  的标准, 在  $K_1, K_2, K_3$  三个时刻用某种测量方法求出读数, 得到表 46.4-20。

表 46.4-20 数据(读数)及其安排  
 $A_1$  的情形

No.	D	E	F	G	数据			计
					$K_1$	$K_2$	$K_3$	
1	1	1	1	1	0.2	-2.9	-4.8	-7.5
2	1	2	2	2	2.1	1.7	1.3	5.1
3	1	3	3	3	1.2	3.0	5.2	9.4
4	2	1	2	3	0.9	-2.1	-4.5	-5.7
5	2	2	3	1	2.0	2.7	2.4	6.6
6	2	3	1	2	0.1	-0.1	0.4	0.4
7	3	1	3	2	0.4	-1.8	-3.2	-4.6
8	3	2	1	3	1.8	-1.3	-3.8	3.3
9	3	3	2	1	0.2	-0.9	0.2	-0.5
计					8.9	-2.2	-6.8	-0.1

$A_2$  的情形

No.	D	E	F	G	数据			计
					$K_1$	$K_2$	$K_3$	
1	1	1	1	1	-0.5	-1.5	2.3	-4.3
2	1	2	2	2	1.3	1.2	0.7	3.2
3	1	3	3	3	0.9	1.4	1.4	3.7
4	2	1	2	3	0.2	-0.3	-0.7	-0.8
5	2	2	3	1	2.8	2.3	1.5	6.6
6	2	3	1	2	1.1	1.0	1.3	3.4
7	3	1	3	2	-0.3	-1.2	-1.8	-3.3
8	3	2	1	3	-0.2	-0.5	-0.7	-1.4
9	3	3	2	1	-0.7	-0.6	-0.3	-1.6
计					4.6	1.8	-0.9	5.5

9.3 数据分析

对于  $A_1$  标准进行计算,

$$S_T = 0.2^2 + (-2.9)^2 + (-4.8)^2 + \dots + 0.2^2 - \frac{(-0.1)^2}{27} = 152.11 \quad (f=26)$$

$$S_{K_1} = \frac{(-K_1 + K_3)^2}{2 \times 9} = 13.09 \quad (f=1)$$

$$S_{K_2} = \frac{(K_1 - 2K_2 + K_3)^2}{6 \times 9} = 0.78 \quad (f=1)$$

$$S_R = \frac{(-7.5)^2 + 5.1^2 + \dots + (-0.5)^2}{3} - \frac{(-0.1)^2}{27} = 93.04 \quad (f=8)$$

时间的变化在各种保存条件下其差异的交互作用

$S_{R \times K_i}$

$$S_{R \times K_1} = \frac{1}{2} [(-0.2 - 4.8)^2 + (-2.1 + 1.3)^2 + \dots + (0.2 + 0.2)^2] - \frac{(-8.9 - 6.8)^2}{18} = 43.99 \quad (f=8)$$

误差变动  $S_e$  为

$$S_e = S_T - (S_{K_1} + S_{K_2} + S_R + S_{R \times K_1}) = 152.11 - (13.69 + 0.78 + 93.04 + 43.99) = 0.61 \quad (f=8)$$

同样对  $A_2$  进行计算后, 作出方差分析表 46.4-21。

表 46.4-21 方差分析表  
 $A_1$  的情形

来源	f	S	V	S'	$\rho(\%)$	
K	R	8	93.04	11.63	92.432	60.77
	L	1	13.69	13.69	13.614	6.95
	q	1	0.78	0.78	0.704	0.46
$K_i \times R$	8	43.99	5.50	43.382	28.52	
e	8	0.61	0.076	1.978	1.30	
T	26	152.11		152.110	100.00	

$A_2$  的情形

来源	f	S	V	S'	$\rho(\%)$	
K	R	8	38.74	4.592	36.804	88.39
	L	1	1.68	1.680	1.663	4.02
	q	1	0.00	—	—	—
$K_i \times R$	8	2.84	0.355	2.704	6.53	
e	8	0.15	0.019	—	—	
(e)	(9)	(0.15)	(0.017)	0.439	0.06	
T	26	41.41		41.410	100.00	

9.4 变化量的预测

假定最初值为  $m$  的物品, 在三个月的变化量为  $\beta$ ,

即

初始  $m$

三个月后  $m + \beta$

六个月后  $m + 2\beta$

表示这种变化偏离初始值  $m$  大小的方差  $\sigma^2$  为

$$\sigma^2 = \frac{1}{3} [(m-m)^2 + (m+\beta-m)^2 +$$

$$(m+2\beta-m)^2] = \frac{5}{3} \beta^2$$

方差分析中用平均值  $\frac{1}{3} [m + (m + \beta) + (m + 2\beta)]$

$= m + \beta$  来求解偏离该值的平均方差  $\sigma_k^2$

$$\sigma_k^2 = \frac{1}{2} [(m - m - \beta)^2 + (m + \beta - m - \beta)^2 +$$

$$(m + 2\beta - m - \beta)^2] = \beta^2$$

把标准延用到  $0, 1, 2, \dots, k$  时

偏离初始值的均方差为

$$\frac{1}{k+1} (0^2 + \beta^2 + 2^2 \beta^2 + \dots + k^2 \beta^2) = \frac{k(2k+1)}{6} \beta^2$$

标准的稳定性偏差

$$\sigma^2 = \frac{k(2k+1)}{6} \times \frac{1}{rS} \times \text{纯变动}$$

### 参 考 文 献

- 1 实验设计法,田口玄一著.魏锡禄,王世芳译.北京:机械工业出版社,1987
- 2 测量技术的实验设计法.田口玄一著.郭玉伟,李静译.北京:机械工业出版社,1988
- 3 试验设计的技术与方法.栾军编著.上海:上海交通大学出版社,1987
- 4 试验设计方法,奥野忠一,芳贺敏郎著.牛长山,张永照译.北京:机械工业出版社,1985
- 5 实验设计与数据分析.俞忠原著.哈尔滨:哈尔滨船舶工程学院出版社,1991

# 第 47 篇 机械系统概念设计

主 编 邹慧君

# 第 1 章 概 论

## 1 机械系统的基本概念

### 1.1 什么是系统

系统这一概念来源于人类长期的社会实践。现代科学技术对于系统思想的发展是有重大贡献的。系统思想是进行分析和综合的辩证思维工具,它在辩证唯物主义那里吸取了丰富的哲学思想,在运筹学、控制论、各门工程学和社会科学那里获得了定性与定量相结合的科学方法,并通过系统工程充实了丰富的实践内容。

如果我们撇开一切具体系统的形态和性质,我们可将系统定义为:具有特定功能的、相互间具有有机联系的要素所构成的一个整体。在美国的韦氏(Webster)大辞典中,“系统”一词被解说为“有组织的或被组织化的整体;结合着的整体所形成的各种概念和原理的综合;由有规则的相互作用、相互依存的形式组成的诸要素集合等等”。在日本的 JIS 标准中,“系统”被定义为“许多组成要素保持有机的秩序,向同一目的行动的集合体”。一般系统的创始人 L. V. 贝塔郎菲(L. V. Bertalanffy)把“系统”定义为“相互作用的诸要素的综合体”。美国著名学者阿柯夫(Ackoff R. L.)认为:系统是由两个或两个以上相互联系的任何种类的要素所构成的集合。

综上所述,一个形成系统的诸要素的集合永远具有一定的固有特性,或者表现为一定的行为,而这些特性或行为是它的任何一个部分都不具备的。一个系统是一个由许多要素所构成的整体,但从系统功能来看,它又是一个不可分割的整体,如果硬把一个系统分割开来,那么它将失去其原来的性质。在物质世界中,一个系统中的任何部分可以被看作为一个子系统,而每一个系统又可以成为一个更大规模系统中的一部分。这就体现了分析与综合有机结合的思想方法。

系统是由要素组成的,离开了要素就谈不上系统。要素是系统的最基本的成分,因此,要素也就是系统存在的基础。一般地说,系统的性质,是由要素决定的,有什么样的要素,就有什么样的系统。要素在构成系统、决定系统时,各种要素要形成一定的结构。要素以一定的结构形成系统时,各种要素在系统中的地位和作用不尽相同。有些要素处于中心地位,支配和决定整个系统的行为,这就是中心要素;还有一些要素处于非

中心、被支配的地位,称之为非中心要素。系统的性质取决于要素的结构,结构的好坏是由要素之间的协调作用直接体现出来的。优质的要素如果协调得不好,形成的结构可能不是最优的;但是,质量差一些的要素,如果协调得好,则可能形成优异的结构,从而决定出质量较优的系统。因此,处理好要素与要素、要素与系统的关系,对于系统的功能和性质至关重要。这就体现出系统设计的重要意义。

系统与环境同样也存在着密切的关系和联系。每一具体的系统都是在时空上有限的存在。作为一个有限的存在,都有它外界的存在或环境。一般把一个系统之外的所有其他事物或存在,称为该系统的环境。环境是系统存在的外部条件。环境对系统的性质起着一定的支配作用。系统的整体性是在系统与环境的相互联系中体现出来的。系统和它的环境构成一个整体。

### 1.2 什么是机械系统

机械系统的关键部分是机械运动的装置,它用来完成一定的工作过程,以代替人类的劳动。现代机器通常由控制系统、信息测量和处理系统、动力系统以及传动和执行机构系统等组成。现代机器中控制和信息处理由电子计算机来完成。不管现代机器如何先进,机器与其他装置的主要不同点是产生确定的机械运动,完成有用的工作过程。因此,实现机械运动的传动和执行机构系统是机械的核心,机器中各个机构通过有序的运动和动力传递最终实现功能变化、完成所需的工作过程。从运动的角度来说,机器中的运动单元体称为机构。因此,机构是把一个或几个构件的运动,变换成其他构件所需的具有确定运动的构件系统。从现代机器发展趋势来看,机构中的各构件可以都是刚性构件,也可以某些构件是柔性构件、弹性构件、液体、气体和电磁体等。而且将各驱动元件与执行系统配合在一起用。

机械是机构和机器的总称。

此外,在实际生产过程中,还有采用多种机器组合起来、完成比较复杂的工作过程,这种机器系统称之为生产线。

从系统的概念来考虑问题,上述构件系统、机构系统和机器系统均可称之为机械系统,只是它们的组成要素各不相同。从完成单一的运动要求考虑,机构就是机械系统,它的组成要素是构件;从完成某一工艺动作过程考虑,机器也是机械系统,它的组成要素是机构;

从完成某一复杂的工艺动作和工作过程考虑,生产线也是机械系统,它的组成要素是机器。如果我们从对某一机器进行加工制造的需要出发,将其中的各个零件作为它的组成要素,因此零件组成的系统也可称为机械系统。由上述分析可见,机械系统是一个广义的概念,它的内涵要按分析研究的对象来加以具体化。

我们可以从广义上来定义机械系统,机械系统是由各个机械基本要素组成的,完成所需的动作(或动作过程)、实现机械能变化、代替人类劳动的系统。机械系统的特点是必须完成动作传递和变换、机械能的利用。这是机械系统区别于其他系统的关键所在。

由于动作的实现方式和完成的具体功能不同,机械系统的种类形形色色。例如,液压系统、气动系统、物流输送系统、自动加工系统等等,均是机械系统。

机器的种类繁多,结构也愈来愈复杂。但从实现机器功能的需要来看,一般应该包括下列子系统:动力系统、传动—执行系统、操纵系统及控制系统等。这些子系统分别实现各自的分功能,综合实现机器的总功能。从完成机器的工作过程需要考虑,传动—执行系统是机器的核心。因此,一般情况下,机械系统研究的重点也是传动—执行系统。研究机械系统概念设计时把重点放在传动—执行机构系统上,其依据是显而易见的。

从系统设计的角度来看,把机械系统界定在机器是比较合理的,有利于开展机器的创新设计。现在有不少文献和专著中把机构也称之为机械系统,从系统的观点来看这是正确的,但是对机构的结构、运动学和动力学的研究在机构学中已经有了深入和全面的展开,也是机构学的主要研究内容。如果把机构学的研究改称为机械系统的研究,反而易使人产生误解。把机器称之为机械系统,有两方面的作用。一是将机器各组成部分作为组成要素可以按系统科学的方法来研究机器的设计,有利于机器的创新和达到综合最优的目标;二是有利于将机器的内部系统与外部环境的外部系统综合在一起形成一个广义机械系统,使其成为人一机一环境的综合体,由此出发进行机器的设计可以达到满足人机工程要求和适应环境变化的优良水平。

### 1.3 什么是广义机械系统

任何一台机器要达到最有效能的运行均离不开人和环境所构成的外部条件。我们把机器本身称之为内部系统,把人和环境称之为外部系统。内部系统和外部系统组成了全系统,也可称之为广义机械系统,如图 47.1-1 所示。人与环境是机械系统存在的外部条件,人与环境对机械的效能起着一定的支配作用。机械系统的整体性是在内部系统与外部系统的相互联系中体现出来的。例如,一台精密加工机床的效能好坏与操作

者的生理、心理和技术水平有关,也与环境对机床的影响有关。



图 47.1-1

## 2 机械系统的基本特征

### 2.1 整体性

整体性是机械系统所具有的最重要和最基本的特性。系统是由两个或两个以上的可以相互区别的要素构成的统一体。虽然各要素具有各自不同的性能,但它们结合后必须服从整体功能的要求,相互间需协调和适应。一个系统整体功能的实现,并不是也不可能是某个要素单独作用的结果。一个系统的好坏,最终体现在它的效能上。因此,必须从整体效能的优劣来判断系统的好坏。确定各要素的性能和它们间的联系时,必须从整体着眼、从全局出发。并不要求所有要素都具有完美的性能。所有要素的性能都十全十美,其整体效能若统一性和协调性差也得不到令人满意的结果。相反,即使某些要素的性能并不很完善,但如能与其相关要素处于很好的统一与协调之中,往往也可使系统具有令人满意的效能。整体性也就是统一性和协调性。

各要素的随意组合不能称为系统。因此,系统的整体性还反映在组合成系统的各要素之间的有机联系上。正是这种有机联系,才使各要素组成一个整体,如果失去了这种有机联系也就不存在整个系统。同样,在系统中不存在与其他要素不发生联系的独立要素。由此可见,系统是不能分割的,不能把一个系统分割成相互独立的子系统。但是,实际的系统有时是很复杂的,为了研究方便,可根据需要,按功能分解原理把一个系统分解成若干个子系统。这种将系统“分解”所得的子系统与毫无道理的“分割”所得的系统是完全不同的概念。因为在分解系统时,始终保持着代表某一子功能的子系统之间的有机联系。分解后的子系统都不是完全独立的,而是维持着某种联系。这种联系分别用相应的子系统的输入与输出表示。因此,这种子系统也就不能分割成完全独立的要素。

### 2.2 相关性

组成系统的要素是相互联系、相互作用的,这就是

系统的相关性。相关性就是系统各要素之间的特定关系。其中包括系统的输入与输出的关系,各要素间的层次关系,各要素的性能与系统整体之间的特定关系等。系统的相关性还体现在某一要素的改变将影响其对相关要素的作用,由此对整个系统产生影响。

系统的相关性是通过相互联系的方式来实现的,例如有时间的联系和空间的联系。广义的讲,要素之间一切联系方式的总和,叫做系统的结构。不同的联系方式对系统的相关性有不同的影响和作用。没有按一定的结构框架组织起来的多要素集合是一种非系统。结构不能离开要素而单独存在,只有通过要素间相互作用才能体现其客观存在。要素和结构是构成系统的两个缺一不可的方面,系统是要素与结构的统一。给定要素和结构两方面,才算给定一个系统。系统的相关性就是通过结构来体现的。

### 2.3 层次性

系统作为一个相互作用的诸要素的总体,它可以分解为一系列的子系统,并存在一定的层次结构,这是系统空间结构的特定形式。在系统层次结构中表述了在不同层次子系统之间的从属关系或相互作用关系。在不同的层次结构存在着动态的信息流和物质流,构成了系统的运动特性,为深入研究系统层次之间的控制与细节功能提供了条件。

从机械系统的构成来看,由基本要素到系统整体是有阶梯性和层次性的。每个层次反映了系统某种功能实现方式。层次本身就是系统构成的部分。图 47.1-2 表示机械系统的层次结构。

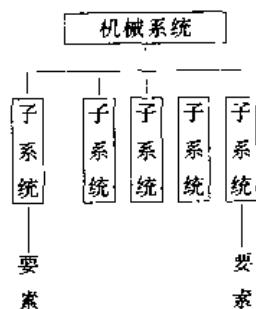


图 47.1-2

如何划分层次、层次的基本特性是什么?只有根据某一具体的机械系统来加以考虑,而且还与系统的分析和设计人员的某些构想有关。

### 2.4 目的性

系统的价值体现在实现的功能上,完成特定的功能是系统存在的目的。系统的目的性是区别这一系统和那一系统的标志。系统的目的性一般用更具体的目标来体现,一般来说,比较复杂的系统都具有不止一个的

目标,因此往往需要一个指标体系来描述系统的目标。

在指标体系中,各个指标之间有时是相互矛盾,有时是互为消长的。为此,要从整体要求出发力求取得全局综合最优的效果,要设法在矛盾的目标之间做好协调工作,寻求平衡点,取得综合最优的方案。

系统的功能就是系统的目的性,它主要取决于要素、结构、环境。要素必须具备必要的性能,否则难以达到预期的目的。要素的相互联系方式取决于系统的结构,选择最佳的结构框架,将有利于最优实现系统的目的。同时,还要选择或创造适当的环境条件,使环境条件有利于系统功能的实现。在要素和环境条件已经给定的情况下,系统的结构才是起决定性影响的。

为了实现系统的目的,系统必须具有控制、调节和管理的功能,这些功能使系统进入与它的目的相适应的状态,实现要求的功能并能排除或减少有意的干扰。

### 2.5 环境适应性

任何一个系统都存在于一定的物质世界的环境中。因此,它必然也要与外界环境产生物质的、能量的和信息的交换,外界环境的变化必然会引起系统内部各要素之间输出、输入的变化,从而会使系统的输入发生变化,甚至产生干扰引起系统功能的变化。不能适应外部环境变化的系统是没有生命力的,而能够经常与外部环境保持最优适应状态的系统,才是理想的系统。

外部环境总是不断变化着的,系统也总是处于动态过程中,稳态过程是相对的、暂时的。因此,为了使系统运行状态良好,必须使系统对外部环境的各种动态变化和干扰具有良好的动态适应性。

为了把握好系统,必须了解系统所处的环境,分析环境对系统有何影响,如何使系统适应这种影响。系统与环境的相互作用、相互联系是通过交换物质、能量、信息来实现的。研究系统和环境的物质、能量、信息交换的规律和特性,才能有的放矢解决系统的环境适应问题。

## 3 机器的类别和基本特征

### 3.1 机器的类别

机械系统的概念是广泛的,但是从机械产品设计需要出发,我们重点研究机器的概念设计。因此,我们应对机器的类别作比较系统的研究。

机器的种类繁多,形形色色,但是从它们的工作类型来分,机器可以分为三类:动力机器、工作机器和信息机器。

动力机器的功用是将任何一种能量变换成机械能,或将机械能变换成其他形式的能量。例如,内燃机、

压气机、涡轮机、电动机、发电机等都属于动力机器。

工作机器的功用是完成有用的机械功或搬运物品。例如,金属切削机床、轧钢机、织布机、包装机、汽车、机车、飞机、起重机、输送机都属于工作机器。

信息机器的功用是完成信息的传递和变换。例如,复印机、打印机、绘图机、传真机、照相机等都属于信息机器。

不管现代机器如何先进,机器与其他装置的主要不同点是产生确定的机械运动,完成有用的工作过程,随之也发生能量的变换。不论是动力机器、工作机器还是信息机器,虽然它们的工作原理各不相同,但是任何机器都必须产生有序的运动和动力传递,并最终实现功和能的变换,完成特有的工作过程。有序运动和动力的传递,主要是依靠机器的运动系统,也就是传动·执行机构系统。因此,机械运动方案设计就成为机器设计的一个关键。

按机器的工作类型来划分机器,可以将众多的机器分成三种机器类型,这将有利于寻找机器设计的一般规律,根据机器的工作特点来进行机械运动系统的创新设计。

### 3.2 能量流、物质流和信息流

机械系统与其他系统一样都存在着能量流、物质流和信息流的传递和变换。机械系统的能量流、物质流和信息流又有它们特殊的形态和变化规律。

#### (1) 能量流

能量流在机械系统中存在于能量变换和传递的整个过程之中。它是机械系统完成特定工作过程所需的能量形态变化和实现动作过程所需的动力。没有能量流也就不存在机械系统的工作过程。在机械系统中能量流又有其特定的变化规则,亦即机械系统中存在机械能转换成其他形态的能,或者其他形态的能转换成机械能。机械能的互换是机械系统主要的能量流特征,没有这种转换也就不能成为机械系统。

能量的类型也是多种多样的。例如,机械能、热能、电能、光能、化学能、太阳能、核能、生物能等等。机械系统的动能和位能均属于机械能。

电动机是将电能转换成机械能。内燃机是将燃油的化学能通过燃烧变成热能,再由热能变成机械能。发电机是将机械能转换成电能。压气机把机械能转换成气体的位能等等。

#### (2) 物质流

物质流在机械系统中存在的主要形式是物流,它是机械系统完成特定工作过程中工作的对象和载体。没有物流也就体现不出机械系统的工作过程和工作特点。

物料的种类也是多种多样的。例如,金属材料包括黑色金属和有色金属材料,纺织品包括麻、棉、丝等,塑料包括容器、薄膜等,还有皮革、橡胶,各种液体,各种气体等。

物流是物料的运动形态变化、物料的构型变化以及两种以上物料包容和混合等的物料变化过程。机械系统的物料只有形态、构型、包容、混合的变化,也就是物料只产生物理的、机械的变化。

机械运动系统所实现的工艺动作过程就是为了满足特定的物料变化过程。

金属切削机床是将金属毛坯通过上料、切削、下料来得到所需形态的某种零件。织布机械是将纱线织成布匹。包装机械是将物件包入包装容器。汽车是将人或货物运送到规定场所。挖土机是将土壤挖开、运送被挖土块。

#### (3) 信息流

信息流是反映信号、数据的检测、传输、变换和显示的过程。信息流的功用是实现机械系统工作过程的操纵、控制以及对某些信息实现传输、变换和显示。因此,信息流对于机械系统实现有序、有效的工作过程是必不可少的。

信息的种类是多种多样的。例如,某些物理量信号,机械运动状态参数,图形显示,数据传输等等。

在工作机器中信息流是实现机械系统的操纵和控制必不可少的。例如,加工中心的工作过程完全是根据给定的信息、数据来控制的。

在信息机器中,信息流的作用更加突出。例如,照相机根据所拍摄景像的远近、外界光线的强弱确定距离和光圈大小,以及曝光时间。通过成像原理获得清晰的景像。

从上述分析可见,任何一台机器的主要特征是从能量流、物质流和信息流中体现出来的,要设计一台新机器首先应从剖析能量流、物质流和信息流着手。

### 3.3 机器的基本特征

任何机器从总体上看都是实现某种能量流、物质流和信息流的传递和变换的。如图 47.1-3 所示。因此,可以这样说,任何一种机器都是实现输入的能量、物料、信息和输出的能量、物料、信息的函数关系的机械装置。新机器的设计就是为了建立实现这种函数关系的机械系统。

通过对输入的能量、物料、信息形态和输出的能量、物料、信息形态的深入分析,可以求出机器所要实现的功能。再通过功能分析和功能求解来构思和设计新机器的运动方案。

分析三类机器的基本特征将会找到一些设计新机



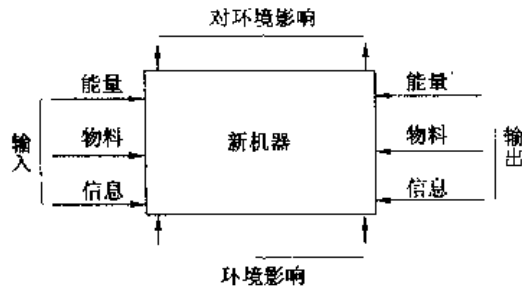


图 47.1.3

器的线索，便于构思和设计。

1. 三类机器的基本特征(见表 47.1-1)

由表 47.1-1 可见，动力机器的最基本特征是其其他

形式能量转换成机械能、或将机械能转换成其他形式的能量。这种能量变换就是动力器的主要功能。

同样，从表 47.1-1 可见，工作机器是利用机械能来搬移物料或改变物料的构形。因此，它最基本的特征是在物料产生运动、改变构形、进行包容等。工作机器的动作过程相对比较复杂。

信息机器的主要功能是传递和变换信息。对于普通的印刷机，其传递和变换原理比较简单，经过给纸、匀墨、印刷到收纸等动作完成。对于静电复印机，其工作原理较为复杂，它由控制系统、曝光系统、成像系统以及搓纸、输纸及图像转印系统组成，完成曝光、显影、转印、定影等工作。

(2) 动力机器类别与功能

表 47.1-1 三类机器的基本特征

类别	特征	能量流	物料流	信息流	举 例
动力机器		1. 将其他能量转换成机械能 2. 将机械能转换成其他能量	为了实现能量变换所需的物质运动变换	控制能量变化的速度和大小	1. 内燃机、电动机等 2. 发电机、压气机等
工作机械		1. 实现物料搬移所需的机械能 2. 实现物料形态变化所需的机械能	物料从一位置搬移至另一位置 上料, 切削, 下料(上料, 包装, 下料)	控制物料搬移 控制上料, 加工, 下料	1. 起重机、汽车等 2. 金属切削机床、包装机等
信息机械		实现信息传递和变换时所需的能量。这种能量较小	信息载体的输送和转移	相关信息的传递和变换	绘图机、复印机、照相机等

1) 化学能转换成机械能的动力机器 有汽油机、柴油机、蒸汽轮机、燃气轮机等，它们将油或煤燃烧后，将化学能变成热能形成高压燃气或高压蒸汽由此产生机械能。这种动力机器，关键是如何有效地将化学能变成热能，由热能转换成机械能的机械装置的结构一般不太复杂，这类动力机器的设计较多地涉及到热能学科。

2) 电能转换成机械能的动力机器 有三相异步电动机、直流电动机、变频电动机、伺服电动机、步进电动机等，它们将电能转换成机械能，这类动力机器的设计主要应用电磁理论和电工学。

3) 机械能转换成其他形式能的动力机器 有压气机、水泵、发电机等。这类动力机器的设计需按相关的转换原理，涉及到各个专业的知识。

(3) 工作机器的类别和功能

工作机器的种类繁多，是三种机器中类别最多的一种。过去这类机器往往按行业来分，有机床、重型机械、矿山机械、纺织机械、农业机械、轻工机械、印刷机械、包装机械等等。按行业和用途类型来划分机器类别对生产和应用是有利的。但是从设计的角度看，采用按工作特点来对机器进行分类是比较有利的。比较广

泛应用的机器可以分成如下几种：

1) 金属切削机床 如车床、铣床、刨床、磨床、钻床、镗床、加工中心等。它们主要的工作特点是工件和刀具的夹持和相对运动情况。按物料输入、输出状况可确定机床的类别和组成特点。

2) 运输机械 如起重机、输送机、提升机、自动化立体仓库等。它们的工作特点是搬运物料、堆积货物。按物料类别不同和搬运要求确定机器的类型。

3) 纺织机械 如各种纺织机。它们的工作特点是将纱线按要求进行纺纱、织布，按纺纱和织布的不同工作原理来确定机器的类型。

4) 缝制机械 如各种平缝机、包缝机、绷缝机、钉扣机、锁眼机、绣花机等。它们的工作特点是按缝制要求运送衣料和缝线，形成衣料成品。不同的缝制要求就构成不同的缝制设备。

5) 包装机械 如糖果包装机、啤酒罐装机、软管充填封口机、制袋充填包装机等等。它们的工作特点是将物料(包括固体、液体、汽体)充入容器，或用包装材料包容物料。由于物料形态不同，包装物具体情况相差较大，包装机械的执行动作构想和执行动作配合就会有不同的方案。

工作机器的设计关键在于如何构想物料的动作过程,实现相应的工艺动作过程。

#### (4) 信息机器的类别与功能

信息机器的种类不算多,一般有打印机、复印机、传真机、绘图机、照相机等。信息机器的功能是进行文字、图像、数据等的传递、变换、显示和记录。信息机器由于工作原理的不同,而具体的结构形式也多种多样。信息机器是精密机械、传感技术、计算机控制技术、微电机技术等多种技术融为一体,并且发展成为典型的机电一体化产品。例如,打印机由打印机构、字车机构、走纸机构三部分组成。静电复印机由曝光、控制、成像和搓纸输纸图像转印等四部分组成。绘图机是通过接口接受计算机输出的信息,经过控制电路向 X 轴步进电动机和 Y 轴步进电动机发出绘图指令,由电动机驱动滑臂和笔爪滑架移动。同时逻辑电路控制绘图笔运动,在绘图纸上绘制所需图形。

## 4 机械设计概述

表 47.1-2 产品设计的类型

设计类型	设计的主要特点
开发性设计	在工作原理、结构等完全未知的情况下,应用成熟的科学技术或经过试验证明是可行的新技术,设计出过去没有过的新型机器。这是一种完全创新的设计
适应性设计	在原理方案基本保持不变的前提下,对产品作局部的变更或设计一个新部件,使产品在质和量方面更能满足使用要求
变型设计	在工作原理和功能结构都不变的情况下,变更现有产品的结果配置和尺寸,使之适应于更多的容量的要求。这里的容量含义很广,如功率、转矩、加工对象的尺寸、速比范围等等

机械产品设计中开发性设计目前还占少数,但是随着产品竞争加剧,开发性设计会有所增加。为了充分发挥现有机械产品的潜力,适应性设计和变型设计就显得格外重要。但是,作为一个设计人员,不论从事哪一类设计,都应该在“创新”上下功夫。“创新”是开发性设计、适应性设计和变型设计的灵魂,“创新”可使设计焕然一新。为了使企业在市场竞争中立于不败之地,必须重视各种类型的设计。

### 4.2 机械设计的一般程序

不论哪一类设计,为了提高机械设计的质量,必须有一个科学的设计程序。通常较为广泛实施和应用的程序可归纳成表 47.1-3 所示的框图程序。

#### (1) 产品规划

产品规划要求进行需求分析、市场预测、可行性分析,确定设计参数及制约条件,最后给出详细的设计任务书(或要求表),作为设计、评价和决策的依据。

#### (2) 概念设计

需求是以产品的功能来体现的,功能与产品设计的关系是因果关系。体现同一功能的产品可以有多种

### 4.1 设计的基本概念

设计一词的英语为 Design,它源于拉丁语 Designar,由 De(记下)与 Signare(符号、记号、图形等)两词组成。因此,设计的最初含义是将符号、记号、图形之类记下来的意思。随着经济的发展和科学技术的进步,设计的内涵不断向深度和广度发展,设计的含义愈来愈广泛、深刻和先进。

设计是人类改造自然的基本活动之一,设计是复杂的思维过程,设计过程蕴含着创新和发明的机会。设计本身就是创新,没有创新的设计严格来说不能称之为设计。设计的目的是将预定的目标,经过一系列规划与分析决策,产生一定的信息(文字、数据、图形),形成设计,并通过制造,使设计成为产品,造福人类。

设计由于情况不同可以有三类不同的设计类型,详见表 47.1-2 所示。

多样的工作原理。因此,这一阶段的最终目标就是在功能分析的基础上,通过构想设计理念、创新构思、搜索探求、优化筛选取得较理想的工作原理方案。对于机械产品来说,在功能分析和工作原理确定的基础上进行工艺动作构思和工艺动作分解,初步拟定各执行构件动作相互协调配合的运动循环图,进行机械运动方案的设计(即机构系统的型综合和数综合)等,这就是产品概念设计过程的主要内容。

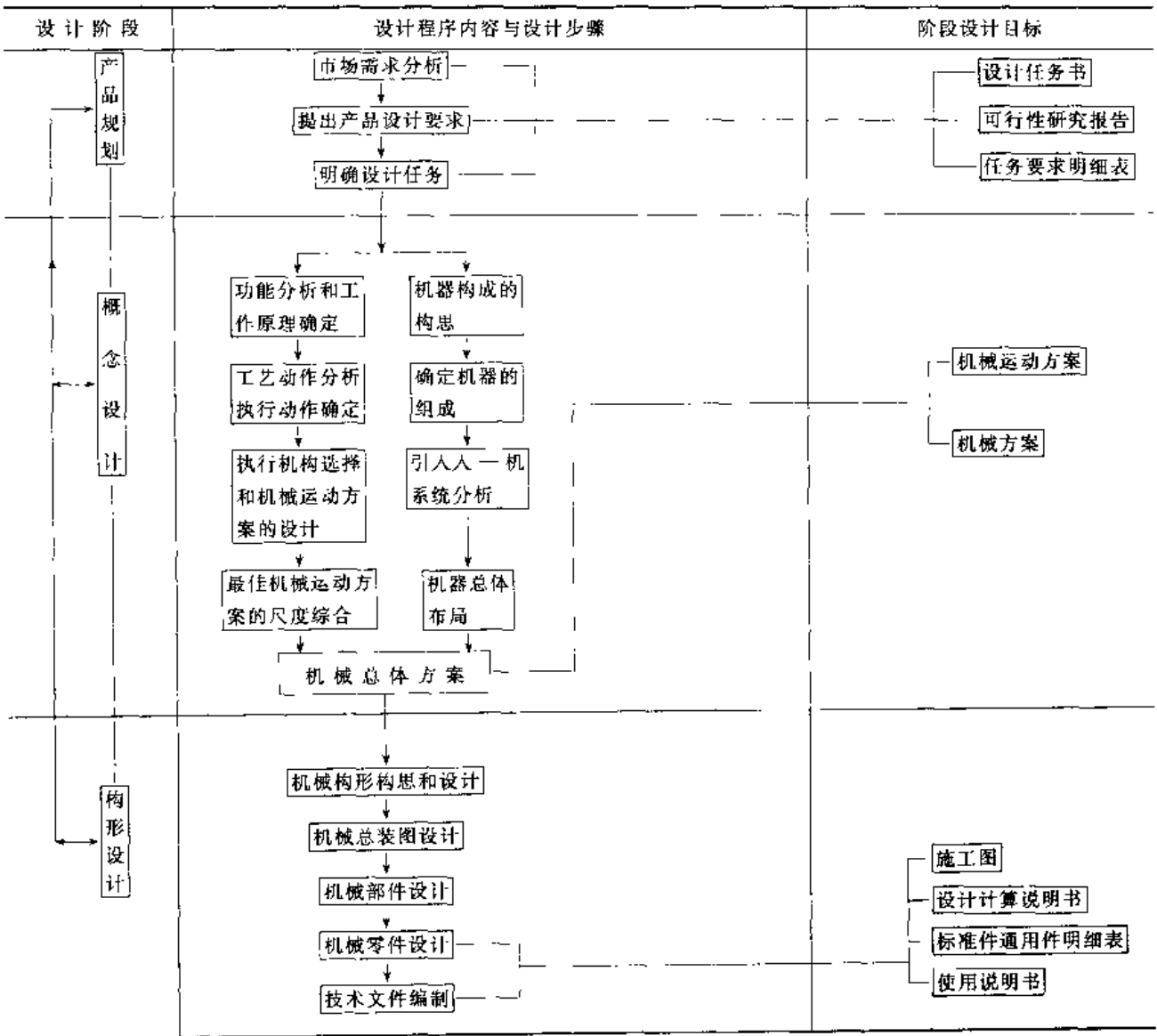
#### (3) 构形设计

构形设计是将机械方案(主要是机械运动方案等)具体转化为机器及其零部件的合理构形。也就是要完成机械产品的总体设计、部件和零件设计,完成全部生产图样并编制设计说明书等有关技术文件。

构形设计时要求零件、部件设计满足机械的功能要求;零件结构形状要便于制造加工;常用零件尽可能标准化、系列化、通用化;总体设计还应满足总功能、人机工程、造型美学、包装和运输等方面的要求。

构形设计时一般先由总装配图分拆成部件、零件草图,经审核无误后,再由零件工作图、部件图绘制出总装图。

表 47.1-3 机械设计的一般程序框图



最后还要编制技术文件，如设计说明书，标准件、外购件明细表，备件、专用工具明细表等等。

## 5 机械系统的概念设计

### 5.1 概念设计与方案设计、创新设计的比较

从机械系统设计的前期工作来看，人们较多提到的名称有方案设计、创新设计和概念设计。

#### (1) 方案设计

在机械设计程序中，方案设计是机械设计的前期工作，它是根据功能要求求出的包括机器各组成部分和功能结构解的机器简图，其中包括结构类型和尺度的示意图及相对关系。这就勾画出机器的方案，可作为构形设计的依据。机器方案设计关键的内容是确定机

器运动方案，通常又称之为机构系统设计方案。

机械运动方案设计必须经过下列步骤：

- 1) 进行机械功能分析；
- 2) 确定各功能元的工作原理；
- 3) 进行工艺动作过程分析，确定一系列执行动作；
- 4) 选择执行机构类型组成机械运动方案；
- 5) 通过综合评价，确定最优机械运动方案。

在方案设计中，设计功能结构，选择功能元的工作原理，进行工艺动作过程分析，确定执行机构类型，组合机械运动方案等设计步骤均孕育着创新。因此，方案设计离不开创新设计，机器创新紧密融合在方案设计之中。

#### (2) 创新设计

重视机器的创新设计是增强机械产品竞争力的根

本途径。那么什么是机器的创新设计?创新设计包含什么内容?

机器创新设计就通过设计人员的创新思维、运用创新设计理论和方法设计出结构新颖、性能优良和高效的新机器。照搬照抄是不可能进行创新设计的。设计本身就应该具有创新。当然,创新设计本身也存在创新多少和水平高低之分。判断创新设计的关键是新颖性,在原理上要新、在结构上要新、在组合方式上要新。

构思一种新的工作原理就可创造出一类新的机器,例如激光技术的应用,产生了激光加工机床。创造一种新结构的执行机构就可造就一种新的机器,例如,抓斗大王包起帆采用了多自由度的差动滑轮组和复式滑轮机构创造发明了“异步抓斗”。采用新的组合方式也可创造出一种新的机器,例如美国阿波罗飞船在没有重新设计和制造一件零部件的情况下,通过选用现有的元器件及零部件组合而成,取得满意的结果,这就是组合的创新。由此可见,创新设计的含义是十分广泛的。

机械的创新设计的内容一般应包括三个方面:

1) 功能解的创新设计 这是属于方案设计范畴的,其中包括新功能的构思、功能分析和功能结构设计、功能的原理解创新、功能元的结构解创新、结构解组成创新等等。从机械方案创新设计角度来看,其中最核心的部分还是机械运动方案的创新和构思。所以,有不少设计人员把机械运动方案创新设计看作机械创新设计的主要内容不是没有道理的。

2) 机械零部件的创新设计 机械方案确定以后,机械的构形设计阶段也有不少内容进行设计创新,如零部件的新构形设计以提高机器工作性能、减小尺寸重量,又如采用新材料以提高零部件的强度、刚度和使用寿命等等,所有这些都是机械创新设计的内容。

3) 工业艺术造型的创新设计 为了增强机械产品的竞争力,我们应该对机械产品的造型、色彩、面饰等进行创新设计。机械的工业艺术造型设计得法,可令使用者心情舒畅、爱不释手,同时也可使机械的功能得到充分的体现。因此,产品艺术造型的创新也是机械创新设计的重要内容。

机械创新设计的内容虽然应包括三个方面,但是最关键的还是方案的创新设计。

### (3) 概念设计

人们对于概念设计的认识和理解还在不断的深入。不管哪一类设计,它的前期工作均可统称为概念设计。例如,很多汽车展览会展示出概念车,它就是用样车的形式体现设计者的设计理念和设计思想、展示汽车设计的方案。又如,一座闻名于世的建筑,它的建筑效果图就体现出建筑师的设计理念和建筑功能表达,

是属于概念设计范畴。

概念设计是设计的前期工作过程,概念设计的结果是产生设计方案。但是,概念设计不只局限于方案设计。概念设计应包括设计人员对设计任务的理解,设计灵感的表达,设计理念的发挥。概念设计还应充分体现设计人员的智慧和经验。因此,概念设计前期工作应充分发挥设计人员的形象思维。概念设计后期工作较多的注意力集中在构思功能结构、选择功能工作原理和确定机械运动方案等,与传统的方案设计没有多大区别。

概念设计由于内涵广泛,使其有更大范围内的创新和发明。

由上述分析可见,概念设计比方案设计更加广泛、深入,因此概念设计包容了方案设计内容。同时,应该看到概念设计的核心是创新设计,概念设计是广义意义上的创新设计。

## 5.2 概念设计的内涵

大家知道 M. French 早在 1974 年编著了一本称之为“工程师用的概念设计”(Conceptual Design for Engineers)的著作,书中对概念设计作了如下的描述:“概念设计就是确定设计任务和用简图形式表达的广义解。概念设计对设计师有很高的要求,要求对产品各种性能有明显的改善。在这一阶段需要将工程科学、专业知识、加工方法以及商业知识等各方面知识融合在一起,以作出最重要的决策。”

Path 和 Beitz 在 1984 年的专著《Engineering Design》(工程设计)中,对概念设计表述为:“在确定任务之后,通过抽象化,拟定功能结构,寻求适当的作用原理及其组合等,确定出基本求解途径,得出求解方案,这一部分设计工作叫作概念设计。”

在 French 和 Path 对概念设计作了表述之后,在 20 多年的时间内,人们对概念设计的研究日益增加、不断深入,使概念设计的内涵更加广泛和深刻。主要体现在:

1) 在设计理念上溶入了设计师以智慧和经验为结晶的崭新的设计哲理和创新灵感,使概念设计更具创新性。

2) 在设计内容上更加广泛,根据产品生命周期各个阶段的要求进行市场需求分析、功能分析、确定功能工作原理、功能载体选择和方案组成等。可见确定方案是概念设计的最终结果,概念设计全过程的好坏才是方案设计的关键。

3) 在设计方法上更加全面地融合各种现代设计方法,寻求全局最优方案,同时使设计过程更具有创造性。

总之,概念设计是方案全面创新的一个设计过程,它集成了设计师的智慧和灵感、先进设计方法的应用、设计资料和数据库广泛采纳、相关的专业知识运用等等。

### 5.3 概念设计的基本特征

#### (1) 创新性

创新是概念设计的灵魂。只有创新才有可能得到结构新颖、性能优良、价格低廉的富有竞争力的机械产品。这里的创新可以是多层次的,如从结构跃进、结构替换的低层次创新工作到工作原理更换、功能结构的修改、整体设计理念的更新等高层次的创新活动都属于概念设计的范畴。

#### (2) 多样性

概念设计的多样性主要体现在其设计步骤的多样化和设计结果的多样化。不同的功能定义、功能分解和工作原理等,会产生完全不同的设计思路和设计方法,从而在功能载体的设计上产生完全不同的解决方案。例如,采用机械传动原理或石英振荡原理就产生机械式手表或石英手表。两种结果完全不同。

#### (3) 层次性

概念设计的层次性体现在两方面。一方面,概念设计分别作用于功能层和载体结构层,并完成由功能层向结构层的映射。如功能定义、功能分解作用于功能层上,而结构修改、结构变异则作用于结构层,由映射关系将两层连接起来。另一方面,在功能层和结构层中也有自身的层次关系。例如功能分解就是将功能从一个层次向下一个层次推进。功能的层次性也就决定了结构的层次性,不同层次的功能对应不同层次的结构。例如结构“自行车”的功能是代步,而自行车的子功能之一“控制行进方向”则是由结构“车把”来完成的。

## 5.4 机械系统概念设计的基本内容

### 5.4.1 功能分析与功能结构设计

1) 功能抽象化:把市场需求和用户要求通过分析,进行功能抽象,突出任务核心和摆脱因循守旧,将会有利于找出新颖的方案;

2) 功能分解:将功能进行分解,使其得到合适的若干子功能,分解过程一定程度上也是创新过程;

3) 功能结构图设计:将各子功能的抽象关系确定后,进行功能结构图的构思和设计。

### 5.4.2 工艺动作的分解和构思

实现机械产品的功能是靠工艺动作来完成的,即一系列工艺动作的目的是完成所需实现的功能。工艺

动作的分解往往对应于功能的分解。例如,缝纫机的缝纫功能分解为刺布、挑线、钩线和送布四大功能,它们所对应的动作为机针上下运动、挑线杆供线和收线、梭子钩线和推送缝料四大动作。又如啤酒灌装机的灌装功能分解为送瓶、灌装、压盖、出瓶四大功能,可用对应的四个动作来完成。同一功能可以由不同的工艺动作实现,因而工艺动作的构思也是相当重要的。例如,在制袋充填封口机中,如直接模拟手工制袋动作,则机构动作非常复杂。但是如果利用相对运动原理逆反思索方法,将制袋成形器采用不动的象鼻形,而塑料薄膜作相对运动,就会使制袋机构大为简化。

### 5.4.3 执行机构系统方案构思与设计

实现功能的工艺动作,在机械系统中是靠若干个执行机构来完成的。机械产品概念设计最终归纳为机械运动方案设计,也就是执行机构系统方案设计。执行机构系统方案的构思与设计是概念设计中非常重要的内容。它的设计内容的框架如图 47.1-4 所示。它分为三部分:动力子系统、传动及执行机构子系统和控制子系统。传动及执行机构子系统是方案设计的核心。其中传动机构和执行机构越来越紧密地连接在一起,而且许多机构同时担负传动和执行的作用,无法分割。因此,在概念设计中将它们作为一个整体对待是合理的。控制机构在机械系统中一般采用各执行机构的主动曲柄的相位差来实现。

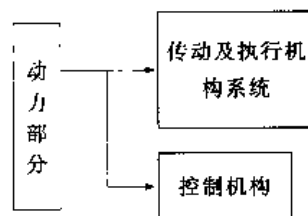


图 47.1-4

## 5.5 机电一体化系统的概念设计

机电一体化系统是现代机械系统,用以实现产品功能和工艺动作过程,它是充分应用电子计算机的信息处理和自动控制功能、可控驱动元件特性的现代化机械系统,体现了机械系统的智能化、自动化。

机电一体化系统从功能分析来看,它的基本构成如图 47.1-5 所示。

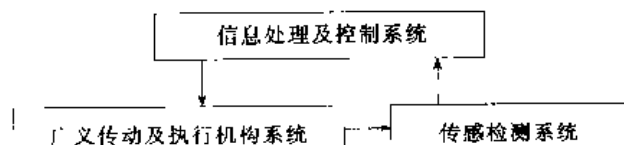


图 47.1-5 机电一体化系统的组成

机电一体化系统主要组成有:广义传动及执行机

构系统,它是通过驱动元件驱动传动及执行机构系统完成工艺动作过程;传感及检测系统进行完成工艺动作过程所需信息的检测,以便有效地实现控制;信息处理及控制系统是由电子计算机及相关软件来实现信息处理和工艺动作过程的控制。

机电一体化系统是现代化的机械系统,从本质上来说还是属于机械系统,因此它的概念设计的基本内容和步骤与机械系统的概念设计没有原则的区别。它

的概念设计过程框架如图 47.1-6 所示。它是从市场需求分析出发,确定机电一体化系统的总功能,通过设计理念的具体构思,进行功能分析和工艺动作过程构思,将工艺动作过程进行分解。选择合适的广义传动、执行机构系统、传感检测系统和信息处理及控制系统。集成机电一体化系统方案,进行若干可行方案的评价,确定最佳的方案。

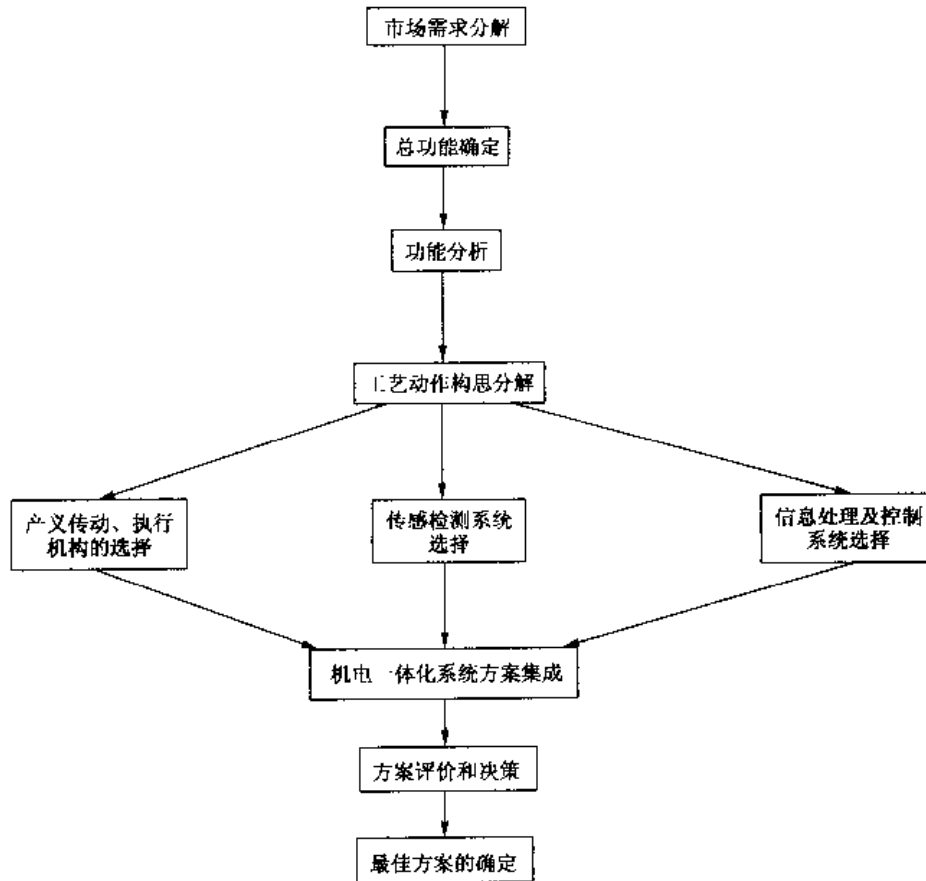


图 47.1.6 机电一体化系统概念设计框架

## 第 2 章 机械系统概念设计的基本方法

### 1 工艺动作过程和执行机构

#### 1.1 工艺动作过程

机械系统概念设计的前期工作就是要进行工艺动作过程的构思。大家知道,机器的功能是通过它的工艺动作过程来完成的。例如,缝纫机的功能是通过刺布—供线—勾线—收线—送布的工艺动作过程来实现缝纫功能。又如,平板印刷机的功能是通过上墨—刷墨—印刷的工艺动作过程来实现印刷功能。再如,扭结式糖果包装机的功能是通过送纸—送糖果—裹包—扭结的工艺过程来实现糖果扭结包装功能。

工艺动作过程取决于所需实现的功能的工作原理,不同的工作原理就会有不同的工艺动作过程。例如,滚齿原理和插齿原理二者的工艺动作过程是不同的。但是采用同一工作原理的工艺动作过程也是可以有不同的工艺动作过程。例如,折叠式包装机,就可以有如图 47.2-1a、b 所示的二种工艺动作过程。

工艺动作过程是实现机器功能所需的一系列动作型式、按一定顺序组合而成的系列动作。它往往可以按一定规则加以分解。

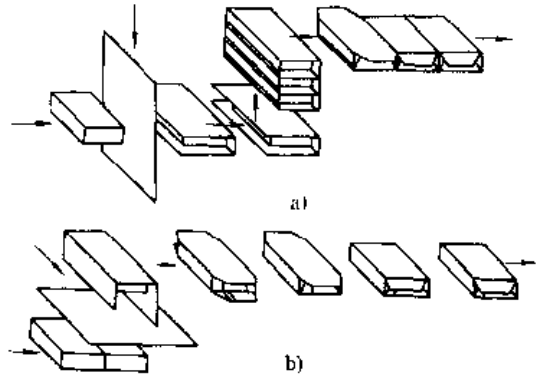


图 47.2-1 阶梯型工艺路线

#### 1.2 执行动作

机器的工艺动作过程一般来说是比较复杂的,往往难以用某一简单的机构来实现。因此,从设计机械运动方案需要出发,我们把此工艺动作过程分解成以一定时间序列表达的若干个工艺动作。这些工艺动作,从机械设计的角度来看称之为机械的执行动作。简称为执行动作。

执行动作五花八门,但是我们可以归纳为表 47.2-1 所示的八种类别。

表 47.2-1 执行动作的类别

序号	执行动作的类别	具体说明
1	连续旋转运动	包括等速旋转运动、不等速旋转运动
2	间歇旋转运动	实现不同间歇要求的间歇旋转运动
3	往复摆动	实现不同摆角的往复摆动
4	间歇往复摆动	实现不同间歇停顿的来回摆动
5	往复移动	实现不同行程大小的往复移动
6	间歇往复移动	实现不同间歇停顿的往复移动
7	刚体导引	实现连杆型构件的若干位姿
8	预期运动轨迹	实现连杆上某些点的给定轨迹

#### 1.3 执行构件和执行机构

机械中完成执行动作的构件,称为执行构件。一般情况下,机构的从动件不止一个,从动件中执行构件至少有一个。执行构件是实现预期执行动作的从动件。执行构件也称为输出构件。在某些场合,执行动作同时由机构的两个执行构件来完成。为了有利于完成执行动作,执行构件往往需要做成特殊构形。

实现各执行构件所需执行运动的机构,称为执行

机构。一般来说,一个执行动作由一个执行机构来完成,但也有用多个执行机构完成一个执行动作,或用一个执行机构完成多个以上的执行动作。

在机械系统运动方案的确定过程中,执行动作的多少、执行动作的形式以及它们间的协调配合等都与机械的工作原理、工艺动作过程及其分解等有密切关系。

确定执行动作、选择执行机构是机械系统运动方案设计中富有创造性的设计内容。

## 2 工作原理和工艺动作分析

### 2.1 机械工作原理的确定

实现机械功能的方法虽然都局限于机械学范畴，但是还是可以根据不同的情况来选择不同工作原理，构思各种动作形式。具体举例如下：

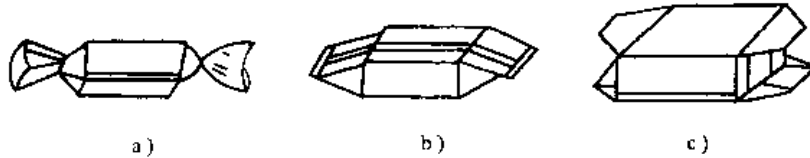


图 47.2-2 颗粒糖果包装形式

3) 折叠式 见图 47.2-2c，它将颗粒糖果用纸进行侧封、端封后实现糖果包装。

三种不同的包装原理，它们的工艺动作过程和工序各不相同，就需选择不同的执行机构，构成相应的运动方案。

#### 2.1.2 印刷工作

根据印刷数量的大小、质量要求高低，印刷机的工作原理有二种形式。

1) 平板印刷工作原理 是在平面的铜锌字版上涂刷油墨，然后将待印纸品覆盖在铜锌字版上完成印刷工艺。平板印刷工作犹如盖图章动作；

2) 轮转式印刷工作原理 将铜锌字版装在圆筒表面，随圆筒连续转动，而将待印纸张作相应连续移动并与圆筒贴合，完成印刷工艺。轮转式印刷的工作原理与盖章式的动作不同，它是属于连续式的，因此生产效率大大提高。

平板印刷机和轮转式印刷机，它们的工艺动作过程和工序各不相同，因此它们的机械系统运动方案明显有区别。

#### 2.1.3 螺栓的螺纹加工

螺纹加工的工作原理主要有四种：

1) 车削 用螺纹车刀在车床上切削加工；

2) 套螺纹 用螺纹板牙在机床上(或用手工)加工；

3) 滚压 用一副滚丝轮在滚丝机上滚压加工；

4) 搓丝 用一副搓丝板在搓丝机上加工。

由于四种螺纹加工的工作原理各不相同，它们的工艺动作过程和工序也就各不相同。因此它们的机械运动方案有很大不同，构成运动方案的机构型式也有很大差别。

由上述分析可见，机械系统的工作原理体现了实

#### 2.1.1 糖果包装

根据市场需求和实际需要，糖果包装有三种形式：

1) 扭结式 见图 47.2-2a，它是将颗粒糖果用纸裹包后扭结而成；

2) 接缝式 见图 47.2-2b，又称枕式包装，它是将纸裹包在糖果上进行纵封、横封后得到的糖果包装；

现机械系统功能各种具体可行的方法。它们的具体实现是依靠某些工艺动作过程。

### 2.2 工艺动作过程的分解

图 47.2-3 表示机械系统运动方案求解的过程模型。由此可见，机械系统工作原理确定之后，构思工艺动作过程和对工艺动作过程的分解是机械系统运动方案设计的重要步骤。

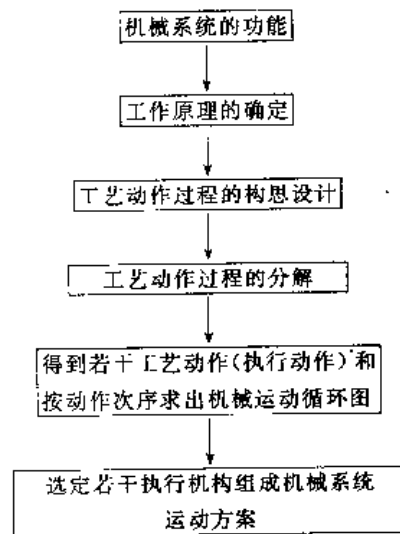


图 47.2-3 机械系统运动方案求解的过程模型

工艺动作过程构思是由机械系统的功能出发，根据工作原理提出可能的动作过程。例如，图 47.2-1 所示的折叠式包装，除了图 a、b 所示的包装动作过程外，还可以有其他的包装动作过程。如前侧先包或下面先包等动作过程。

再如，图 47.2-4 所示冲压式蜂窝煤成型机示意图，它的工艺动作过程可由六个动作组成：加料—冲压成型—脱模—扫屑—转位—成品输送。构思工艺动作过程要满足工作可靠、工序合理、工效提高等多种要



求。

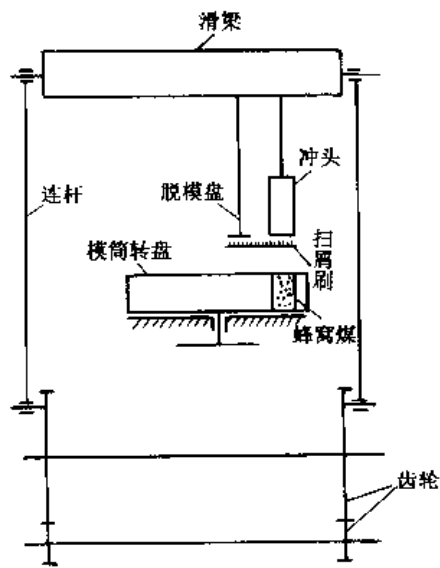


图 47.2-4 冲压式蜂窝煤成型机示意图

又如,图 47.2-5 所示 JB 型缝纫机的示意图,根据衣料缝制要求和缝制工作原理构思出缝制动作过程:刺布—挑线(供线和挑线)—勾线—送布。

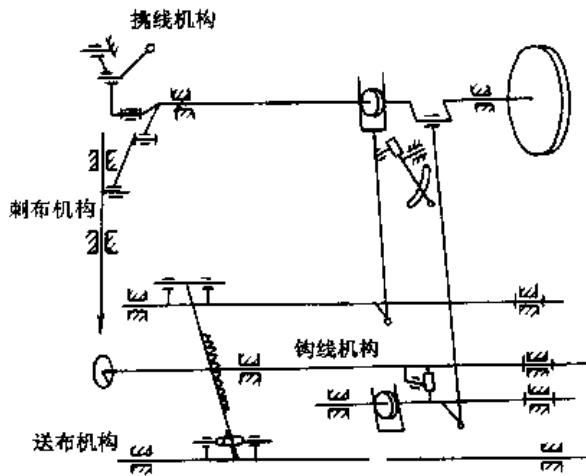


图 47.2-5 JB 型缝纫机示意图

工艺动作过程分解的目的是确定执行动作的数目以及它们之间的时间序列。原则上每个执行动作型式从执行机构所能完成的八种执行动作中选择,以便从现有机构中选择合适的执行机构。否则需用机构创新设计方法创造新的执行机构,完成特殊的执行动作。

工艺动作过程分解的方法,一般有以下几种:

1) 将构思工艺动作过程进行逆向分析,不难得到各个执行动作。构思工艺动作时往往是将预先考虑的若干执行动作作用时间顺序贯穿在一起的。

2) 用类比法对工艺动作过程进行分解。即借鉴相似工艺动作过程分解的办法来进行动作过程分解不失

为一种好办法。

3) 采用拟人动作方法来实施工艺动作过程分解。如借鉴人手对一物体的包装过程来得到折叠式包装的各个执行动作。

### 3 系统设计方法

#### 3.1 系统设计基本概念

系统设计是将机械产品看作一个技术系统,用系统工程方法对机械系统运动方案进行分析和综合。机械系统运动方案是以一定的工艺动作过程来实现机械系统的功能。机械系统运动方案中实现工艺动作过程各个执行动作的相应执行机构均是系统中的各个子系统,各子系统与系统整体之间是密切相联的,通过各子系统的配合和协调使系统达到机械系统运动方案的最优目标。由此获得最佳的方案设计。

在进行机械系统运动方案设计过程中,为了做好系统设计还应该重视系统分析,系统分析是为了更好地设计。

#### 3.2 系统分析

机械系统运动方案的系统分析是将实现机械系统总功能的机械工艺动作过程通过动作分解求得若干执行动作,选用或设计相应的执行机构。以这些执行机构组成的机构系统来实现机械工艺动作过程。

机械运动方案的最主要的功用是将一系列的输入运动转换为相应的输出运动,这些输出运动构成了机械的工艺动作过程。很明显这些输出运动必须具有有序性和必要的反馈性。有序性就是机械运动系统中各个子系统的动作是有序的、相互制约、又能协调地工作。机械运动方案的工作循环图是机械有序性的具体体现。机械的有序性保证机械工作的有效性和整体性。机械中的反馈性是将某些运动信息进行反馈实现系统中各子系统协调动作和有效控制,以保证系统能精确、安全和有效地工作。从现代机械系统要求来看,没有反馈的系统往往不能成为一个好系统。

##### 3.2.1 系统分析的要素

一般情况下,机械运动方案进行系统分析的要素,有以下四个:

1) 目标 经过分析后确定的目标应是必要的、有根据的、可行的。这是设计机械运动方案的根据,也是分析机械运动方案的出发点。

2) 备用方案 这是达到机械运动方案目标的若干可供采用的方案,以供比较和选择。

3) 指标 这是对可用方案进行分析的依据,是衡

量系统目标的具体标志。对于机械运动方案的指标主要有：运动性能，动力性能，尺寸紧凑性，制造安装的难易程度，设计工作量大小与设计复杂程度，维护使用方便性等等。

4) 模型 这是根据机械运动方案的目标要求，用若干参数或因素体现出对系统本质方面的描述。通常来说，对于某一机械运动方案可以用一个包含多项评价指标和指标评价价值的评价体系来描述这个模型。

模型具有三个主要特征：

- (a) 是实际机械系统合理的抽象和有效的模仿；
- (b) 能表征实际系统的本质属性和主要特征；
- (c) 可以表明主要因素间的相互关系。

对于机械运动方案所采用的抽象模型通常有三种类型：

(a) 概念模型。这是人们应用知识、经验和直觉，在缺乏资料的情况下，通过构想一些资料，建立初始模型，再逐步扩展和完善而形成的。在形式上它们可以是思维的、文字的或描述性的；

(b) 图式模型。这是用少量文字、简明的数字、线条等构成的模型，它能够直观形象地表示出方案的一些本质和特征，如流程图、方框图等；

(c) 机械工作循环模型。这是用机械工作循环图建立起各执行机构的时序关系和运动特征的模型，它能够较为具体地表示方案构成和工作特征。

### 3.2.2 系统分析的程序

1) 机械运动方案总功能的分析与确定。通过市场需求和用户要求的调查分析、发现问题、明确机械运动方案的目标。明确要求，就易于确定机械运动方案的总功能。需求分析时，要求论证它的合理性、可行性和经济性。同时，提出各种可行性方案并进行评价选优。

2) 机械运动方案功能体系的构造。将机械运动方案的总功能通过合理的分解，从而设法构造它的功能体系。由此表达了功能的构成、功能的关系和总功能的实现，同时还得出各分功能、子功能的主要参数。

3) 机械运动方案的评价和选优。确定评价指标体系对机械运动方案可能采用的方案进行评价和选优，得到最为满意的方案。

## 3.3 系统设计

### 3.3.1 系统设计的概念

系统设计是在系统分析的基础上进行的。系统分析为系统设计提供了下列条件：

1) 机械运动方案重新设计的必要性、可能性和可行性；

2) 机械运动方案的目标(总功能)和约束条件；

3) 机械运动方案的框架结构和评价基础；

4) 得出几种有价值的可供进一步改进的机械运动方案。

系统设计的任务就是要充分利用系统分析的结果，设计出最大限度满足系统总功能要求的具体机械运动方案。

### 3.3.2 系统设计的基本原则

机械运动方案的系统设计的基本原则：

1) 追求整体最优。方案整体最优，包括它的内部最优状态和方案最优输出两个含义。前者是指方案各环节的平衡、协调，传输的均衡，最低的耗费、最高的效率等；后者是指方案能有最大成果，如最大生产量等。

整体最优不一定是各子系统(执行机构)的优化。有时对系统的局部作过分改进，还可能反而使系统整体变劣。例如，在机械系统运动方案设计中过分强调某一子系统(执行机构)的输入—输出的运动精度要求，如得不到机械系统在整体上的良好配合，会导致整个系统运动精度下降或不能正常工作。只有从整体出发考虑各子系统的优化，才能产生积极的效用。

2) 抓住主导事件。在确定机械运动方案设计的基本目标时，应不考虑或少考虑小概率事件的影响。考虑主导事件的影响来确定机械运动方案的设计要求。对于机械运动方案设计，主导事件是指某一主要的工艺动作过程和动作要求，这些在设计中应努力加以解决，对于一些不常需要的工艺动作过程和动作要求可以少加考虑，使设计的基本目标尽量简化。

3) 信息分类要适应决策的需要。系统设计往往会涉及很广泛的资料和信息，对于机械运动方案设计就会涉及技术性指标、经济指标、市场需求、国内外类似产品的状况、原材料和配件的供应、设计技术资料等等。为了有利于系统设计的进展，应将各种信息进行分类，把决策有关的信息优先加以考虑，以提高决策的速度和质量。

### 3.3.3 系统设计的过程

机械运动方案设计过程主要包括如图 47.2-6 所示七个步骤。

系统设计的主要任务，就是要得到最优的机械运动方案，而在方案的寻优过程中又充满着综合与分析的交互作用，在系统分析设计的全过程中，分析—综合方法是系统设计的基本方法。这种方法能够合理解决系统的目标—功能—结构—效益之间协调和最优化问题。

针对某一特定系统设计的具体步骤有其本身的特点。具体问题要作具体分析,这是在系统设计中必须牢牢掌握的原则。

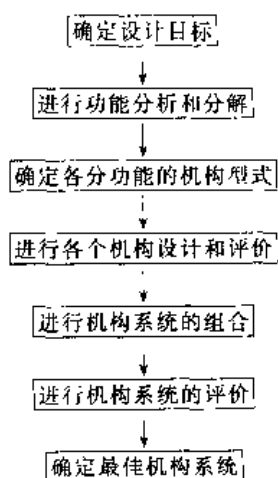


图 47.2-6 系统设计过程

### 3.3.4 系统综合评价

1) 机械运动方案的拟定和设计,最终要求提供最优的方案,而方案的优劣是需要通过系统综合评价来确定的。因此,系统分析和系统设计的基本方法都必须对结果进行评价。从设计全过程来看,评价工作不仅在整个机械运动方案设计完成后是需要的,在整个设计过程的每一阶段也是需要的。

对于某一机械运动方案来说,其目标就是要求完成某一工艺动作过程,亦即完成一系列的动作。对于各个动作的要求有时会相互矛盾。所以,对一个机械运动方案应该建立一个评价体系,通过综合性的评价确定最优的机械运动方案。

2) 综合评价的基本原则大体上有

(a) 要保证评价的客观性。评价的目的是为了决策。因此评价是否客观,就会影响决策是否正确。为了保证评价的客观性,要求评价资料的全面性和可靠性;防止评价人员的倾向性,评价人员组成要有代表性。

(b) 要保证方案的可比性。即要求各个方案在实现基本功能上有可比性和一致性。有的方案在实现个别功能方面优点突出或有新颖独特之处,只能表明它在这方面的优越之处,不能代替它在其他方面所能实现的要求,更不能掩盖它在其他方面存在的不足之处。否则,会失去综合评价的作用,犯“突出一点,不顾其余”的错误。这种主观片面的做法,显然不利于评选最优方案。

(c) 要有适合于方案设计阶段的评价指标体系。评价指标体系是全面反映系统目标要求的一种评价模式。因此,评价体系应该主要考虑机械运动方案总功能

所涉及的各方面要求和指标,不要考虑或少考虑其他方面或设计阶段的要求。建立评价指标体系一定要体现科学性、全面性和专家的经验。

3) 系统综合评价的一般步骤有

(a) 确定系统综合评价的指标体系。对于机械运动方案的评价指标体系一般应包括实现功能、工作功能、动力功能、经济性、结构紧凑等五大类的评价指标,这些大类和具体的评价项目均要与机械运动方案设计内容密切相关。在建立系统评价体系时,应尽可能广泛地听取这一领域内权威专家的意见和建议。

(b) 确定各大类和具体评价指标重要程度的权系数。确定权系数,实际上是使评价指标体系对各种比较特殊的用途和特殊的使用场合的机械运动方案从整体上进行调整。使系统评价指标体系有更大的灵活性、广泛性、实用性,使系统评价指标体系有更大的适用范围。例如,对于重型机械运动方案设计时的评价指标与轻工机械运动方案设计时的评价指标有一定的区别。权系数就可以适应这两者区别的需要。

(c) 对子系统的方案进行逐项评价。得到综合评价指标值,为方案综合评价提供必要的条件。

(d) 对所采用的机械运动方案进行逐项评价。得出各单项评价指标值。

(e) 进行系统单项评价指标的综合。得出评价指标体系各大类的评价值。

(f) 最后进行系统的综合评价。综合机械运动方案各大类指标的评价值,得出整个方案的总评价值。对于多个方案可按总评价值来进行择优,以确定最优方案。在确定最优方案时,还应考虑制造工厂的类似产品情况、加工设备条件、技术力量等等。有时总评价值最高的方案不一定被最后选用,就是由于某一方面因素影响。

图 47.2-7 表示了上述系统综合评价步骤。

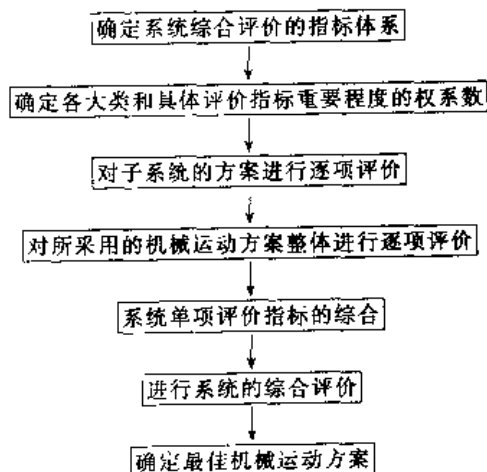


图 47.2-7 系统评价的步骤

## 4 层次分析方法

### 4.1 层次分析法的基本步骤

- 1) 构造机械运动方案的层次结构模型;
- 2) 建立判断矩阵, 计算相对权重, 这又称为层次单排序;
- 3) 判断矩阵一致性检验;
- 4) 计算组合权重, 并通过组合权重的对比, 得到决策方案的优劣顺序, 又称为层次总排序。

### 4.2 层次结构模型

层次结构模型表示方案所涉及的因素及其之间的

目标层 G	实现糖果折叠包装功能				最高层	
准则层 C	实现运动好坏	工作性能	动力性能	经济性	结构紧凑	中间层
方案层 B	方案 B <sub>1</sub>	方案 B <sub>2</sub>	方案 B <sub>3</sub>	方案 B <sub>4</sub>		最低层

图 47.2-9 糖果包装的层次结构

### 4.3 构造判断矩阵和计算相对权重

#### 4.3.1 构造判断矩阵

判断矩阵是将层次结构模型中同一层次的因素相对于上层的某个因素, 相互间作成对比较而形成的矩阵。以图 47.2-10 所示的层次结构为例, 方案层的备选方案 B<sub>1</sub>、B<sub>2</sub>、…、B<sub>n</sub> 相对上层的准则 C<sub>k</sub> 作成对比较, 可构成下面的判断矩阵 P<sub>C<sub>k</sub>-B</sub>, 如图 47.2-10 所示。

图中, b<sub>ij</sub> 是以 C<sub>k</sub> 为准则对 B<sub>i</sub> 与 B<sub>j</sub> 那个更好来确定代表好的程度的数值。

对于 C<sub>k</sub> 为准则(例如工作性能)如何确定图 47.2-10 中元素 b<sub>ij</sub>, 考虑到大多数准则比较往往是模糊的, 哪个方案更好, 或稍差等等, 为了使其定量化, 往往引入判断标度。通常使用 1~9 标度法, 如表 47.2-2 所示。

表 47.2-2 1~9 标度说明

标 度	说 明
1	表示 B <sub>i</sub> 与 B <sub>j</sub> 相比, 两个要素同等好
3	表示 B <sub>i</sub> 比 B <sub>j</sub> 稍微好一些
5	表示 B <sub>i</sub> 比 B <sub>j</sub> 明显好
7	表示 B <sub>i</sub> 比 B <sub>j</sub> 好得多
9	表示 B <sub>i</sub> 比 B <sub>j</sub> 绝对好
2, 4, 6, 8	表示两相邻标度的中间值

通过比较得到的判断矩阵 P=[b<sub>ij</sub>]<sub>n×n</sub> 具有以下性

- 1) b<sub>ij</sub> > 0
- 2) b<sub>ij</sub> = 1/b<sub>ji</sub> } i, j = 1, 2, …, n
- 3) b<sub>ij</sub> = 1/b<sub>ji</sub>

关系。图 47.2-8 所示为递阶层次结构模型。其中最高层通常只包含一个要素, 一般为机械运动方案的总功能; 最低层称为方案层, 通常设置方案的各种备选方案。中间层称为准则层, 列出用来衡量是否达到目标的各项评价准则和评价标准等。

目标层 G	G						最高层	
准则层 C	C <sub>1</sub>	C <sub>2</sub>	……	C <sub>k</sub>	……	C <sub>M-1</sub>	C <sub>M</sub>	中间层
方案层 B	B <sub>1</sub>	B <sub>2</sub>	……	B <sub>i</sub>	……	B <sub>n</sub>		最低层

图 47.2-8 递阶层次结构

例如, 为了确定糖果折叠包装机的机械运动方案, 它的递阶层次结构模型如图 47.2-9 所示。

根据以上特性, 可以证明一个 n 阶的判断矩阵只有 n·(n-1)/2 个元素。对于如图 47.2-8 所示的层次结构, 方案层有 n 个元素、准则层中有 m 个元素。因此可以建立 m 个判断矩阵, 即

$$P_{C_k-B} \quad (k=1, 2, \dots, m)$$

同样道理, 准则层对目标层只有一个判断矩阵, 即 P<sub>G-C</sub> 为 1 个。

所以图 47.2-8 所示的层次结构总共需要构造 m+1 个判断矩阵。

很明显, 在构造判断矩阵时应充分采纳专家的意见, 使之更加符合客观实际, 更具科学性。

同样我们可以看到, 如果机械运动方案比较复杂, 它由好几个执行机构组成, 我们可按每一机构构造相应的判断矩阵, 确定若干个可供采用的执行机构。然后, 再选定若干机械运动方案并在方案整体上构造判断矩阵。

C <sub>k</sub>	B <sub>1</sub>	B <sub>2</sub>	……	B <sub>j</sub>	……	B <sub>n</sub>
B <sub>1</sub>	b <sub>11</sub>	b <sub>12</sub>	……	b <sub>1j</sub>	……	b <sub>1n</sub>
B <sub>2</sub>	b <sub>21</sub>	b <sub>22</sub>	……	b <sub>2j</sub>	……	b <sub>2n</sub>
⋮	⋮	⋮	……	⋮	……	⋮
⋮	⋮	⋮	……	⋮	……	⋮
⋮	⋮	⋮	……	⋮	……	⋮
B <sub>i</sub>	⋮	⋮	……	b <sub>ij</sub>	……	⋮
⋮	⋮	⋮	……	⋮	……	⋮
⋮	⋮	⋮	……	⋮	……	⋮
⋮	⋮	⋮	……	⋮	……	⋮
B <sub>n</sub>	b <sub>n1</sub>	b <sub>n2</sub>	……	b <sub>nj</sub>	……	b <sub>nn</sub>

图 47.2-10 备选方案 B 对准则 C<sub>k</sub> 的判断矩阵 P<sub>C<sub>k</sub>-B</sub>

## 4.3.2 计算权重

权重计算的方法有多种,这里仅介绍二种:

## (1) 方根法

这是一种判断矩阵权重的近似算法,其步骤如下:

首先,计算判断矩阵  $P = [b_{ij}]_{n \times n}$  中每行所有元素的几何平均值,得到向量  $M, M = [m_1 m_2 \cdots m_n]^T$ , 其中

$$m_i = \sqrt[n]{\prod_{j=1}^n b_{ij}} \quad (i = 1, 2, \dots, n) \quad (47.2-1)$$

其次,对列向量作归一化处理,得到相对权重向量  $W, W = [W_1 W_2 \cdots W_n]^T$ , 其中

$$W_i = \frac{m_i}{\sum_{j=1}^n m_j}$$

所谓归一化,是指:

$$(a) 0 \leq W_i \leq 1 \quad (i = 1, 2, \dots, n)$$

$$(b) \sum_{i=1}^n W_i = 1$$

最后,计算判断矩阵  $P$  的最大特征值  $\lambda_{\max}$ , 其近似计算式如下:

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{(PW)_i}{W_i} \quad (47.2-2)$$

其中,  $(PW)_i$  是权重向量  $W$  右乘判断矩阵  $P$  得到的列向量中的第  $i$  个分量。

$\lambda_{\max}$  将用于判断一致性检验。

## (2) 特征向量法

它的计算精度较高。线性代数中,对于实数矩阵  $P = [b_{ij}]_{n \times n}$ , 其特征方程为  $(P - \lambda I)W = 0$ , 特征多项式为  $|P - \lambda I| = 0$ , 其中,  $I$  为单位阵,  $W$  为对应于特征值  $\lambda$  的特征向量。对于特征多项式,经运算可求出  $P$  的  $n$  个特征值  $\lambda_1, \lambda_2, \dots, \lambda_n$ , 而最大特征值是指  $\lambda_{\max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 。另外,称下式为矩阵  $P$  的迹:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = b_{11} + b_{22} + \cdots + b_{nn} \quad (47.2-3)$$

特征向量法计算权重的原理如下:设有  $n$  个物体  $B_1, B_2, \dots, B_n$ , 重量分别为  $W_1, W_2, \dots, W_n$ 。若两两比较物体的重量,其比值可构成  $n \times n$  矩阵  $P$ 。若用重量向量  $W = [W_1 W_2 \cdots W_n]^T$  右乘矩阵  $P$ , 可得下式:

$$PW = \begin{bmatrix} \frac{W_1}{W_1} & \frac{W_2}{W_1} & \cdots & \frac{W_n}{W_1} \\ \frac{W_2}{W_2} & \frac{W_2}{W_2} & \cdots & \frac{W_n}{W_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{W_n}{W_n} & \frac{W_n}{W_2} & \cdots & \frac{W_n}{W_n} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = n \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = nW$$

或  $(P - nI)W = 0$

由矩阵理论可知,  $n$  即为  $P$  的特征值, 且是最大特

征值  $\lambda_{\max}$ ,  $W$  则是对应于最大特征值  $\lambda_{\max}$  的特征向量。

不难看出,特征向量法应首先求出判断矩阵的最大特征值  $\lambda_{\max}$ ; 然后计算对应于  $\lambda_{\max}$  的特征向量  $W$ ; 再对  $W$  作归一化处理, 即得到权重向量。当判断矩阵阶数较高时, 可采用迭代算法编程计算特征值。

## 4.4 判断矩阵的一致性检验

## 4.4.1 完全一致性

根据矩阵理论, 若正互反矩阵  $P = [b_{ij}]_{n \times n}$  对所有的  $i, j = 1, 2, \dots, n$ , 均有  $b_{ij} = b_{ik} / b_{jk}$  成立, 则称  $P$  具有完全一致性。此时正互反矩阵  $P$  具有唯一非零的最大特征值  $\lambda_{\max}$ , 且  $\lambda_{\max} = n$ 。实际上, 由于正互反矩阵的  $b_{ii} = 1 (i = 1, 2, \dots, n)$  且令  $\lambda_{\max} = \lambda_1$ , 由式(47.2-3)可得:

$$\lambda_{\max} - \sum_{i=2}^n \lambda_i = n, \text{ 则 } \sum_{i=2}^n \lambda_i = 0$$

## 4.4.2 一致性检验指标

人们对复杂问题涉及的因素进行两两比较时, 不可能作到判断的完全一致性, 总会存在估计误差。这将导致判断矩阵的特征值和特征向量也带有偏差。设  $P'$  为带有偏差的判断矩阵, 其最大特征值和特征向量设为  $\lambda'_{\max}$  和  $W'$ 。因为  $b_{ii} = 1 (i = 1, 2, \dots, n)$ , 又设  $\lambda'_{\max} = \lambda_1$ , 由式(47.2-3)可得:

$$\lambda'_{\max} + \sum_{i=2}^n \lambda'_i = n$$

通常  $P'$  的  $\lambda'_{\max} \geq n$ , 而  $\lambda'_{\max} - n$  就是除  $\lambda'_{\max}$  以外的其余所有特征值的代数和。与完全一致性相比较:

$$\lambda'_{\max} - n = \sum_{i=2}^n \lambda'_i$$

就表征了  $P'$  的偏差程度, 由此一致性检验指标  $C.I$  构造如下:

$$C.I = \frac{\lambda'_{\max} - n}{n - 1} \quad (47.2-4)$$

由式(47.2-4), 对于任意的判断矩阵, 当  $\lambda_{\max} = n$  时,  $C.I = 0$ , 则判断矩阵具有完全一致性;  $C.I$  的值越大,  $P'$  的估计偏差也就越大, 偏离一致性的程度就越大。

## 4.4.3 随机一致性指标

通常判断矩阵的阶数  $n$  越高, 其估计偏差值随之增大, 一致性也越差, 因此对高阶判断矩阵的检验应当放宽要求。为此引入随机指标  $R.I$  作为修正值, 以

更合理的随机一致性指标  $C.R$  来衡量判断矩阵的一致性。

$$C.R = \frac{C.I}{R.I} \quad (47.2-5)$$

表 47.2-3 随机一致性指标中  $R.I$  的取值

$n$	1	2	3	4	5	6	7	8	9	10
$R.I$	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49

为了说明一致性检验, 举例如下。

例 2-1 选择一个执行机构, 将工作性能、实现运

通常只要  $C.R \leq 0.10$ , 则认为  $P'$  具有满意的一致性, 否则必须重新调整  $P'$  中元素的值。式(47.2-5)中  $R.I$  的值, 可根据判断矩阵的阶数从表 47.2-3 中选取。二阶及以下的判断矩阵总是具有完全一致性。

动、动力性能作为三个评价准则, 现确定各准则的相对重要顺序。现构造下面二个判断矩阵:

$P_a$	选择执行机构	工作性能	实现运动	动力性能	$W$
	工作性能	1	1/3	3	0.258
	实现运动	3	1	5	0.636
	动力性能	1/3	1/5	1	0.106

$$\lambda_{\max} = 4.838; C.R = 0.0371 < 0.1$$

$P_b$	选择执行机构	工作性能	实现运动	动力性能	$W$
	工作性能	1	1/3	3	0.258
	实现运动	3	1	1/5	0.636
	动力性能	1/3	5	1	0.106

$$\lambda_{\max} = 4.838; C.R = 1.54 > 0.1$$

由计算结果可知,  $P_a$  满足要求, 而  $P_b$  偏差太大, 使评价准则排序为: 动力性能、工作性能、实现运动。显然这与人们通常的选择性思维不一致。一致性检验可以帮助发现估计误差对过大的偏差必须加以修正。

### 4.5 层次总排序

层次总排序就是根据层次单排序得到的结果来计算组合权重。然后, 通过比较各要素组合权重的大小, 得到要素的相对重要顺序, 依此确定对备选方案的评价。

对于图 47.2-8 所示的递阶层次结构, 设准则层  $C$  对目标层  $G$  的相对权重列向量为  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_m]^T$ , 方案层  $B$  对  $C$  层各项准则  $C_1, C_2, \dots, C_m$  的权重列向量分别记为  $W_1, W_2, \dots, W_k, \dots, W_m$ , 其中  $W_k = [w_{1k}, w_{2k}, \dots, w_{nk}]^T$  是  $B$  层方案  $B_i (i=1, 2, \dots, n)$  对准则  $C_k (k=1, 2, \dots, m)$  的相对权重列向量。由此构成组合权重计算表 47.2-4, 其中  $\sum$  为  $\sum_{j=1}^m$  的简写。

表 47.2-4 组合权重计算表

$C$	$C_1$	$C_2$	...	$C_m$	组合权重 $V$
	$\alpha_1$	$\alpha_2$	...	$\alpha_m$	
$B_1$	$W_{11}$	$W_{12}$	...	$W_{1m}$	$V_1 = \sum \alpha_j W_{1j}$
$B_2$	$W_{21}$	$W_{22}$	...	$W_{2m}$	$V_2 = \sum \alpha_j W_{2j}$
...	...	...	...	...	...
$B_n$	$W_{n1}$	$W_{n2}$	...	$W_{nm}$	$V_n = \sum \alpha_j W_{nj}$

实际上, 由相对权重列向量  $W_1, W_2, \dots, W_m$  可

构造相对权重矩阵  $W = [W_1, W_2, \dots, W_m]$ , 则组合权重  $V$  可按下式计算

$$V = W \cdot \alpha \quad (47.2-6)$$

### 4.6 层次分析法应用举例

层次分析法对于像构思机械运动方案那样难以建立数学模型, 而又缺乏必要数据的非结构性复杂问题, 在分析决策时十分有效和实用。下面举一实例说明其应用。

例 2-2 为了实现高速间歇运动, 选定了三种执行机构方案:  $I_1$ —圆柱凸轮间歇运动机构、 $I_2$ —槽轮机构、 $I_3$ —棘轮机构。评价准则有  $C_1$ —运动性能、 $C_2$ —动力性能、 $C_3$ —制造难度。试采用层次分析法来选择方案。

解:

1) 建立层次结构模型。

层次结构模型如图 47.2-11 所示。

$G$ 层	高速间歇运动		
$C$ 层	$C_1$ 运动性能	$C_2$ 动力性能	$C_3$ 制造难度
$I$ 层	$I_1$ 圆柱凸轮间歇运动	$I_2$ 槽轮机构	$I_3$ 棘轮机构

图 47.2-11 高速间歇运动机构层次结构模型

2) 建立判断矩阵, 计算相对权重。

判断矩阵有  $P_{G-C}, P_{C_1-I}, P_{C_2-I}, P_{C_3-I}$  等 4 个, 分别是  $G$  对  $C$  层, 以及  $C_1, C_2, C_3$  对  $I$  层的判断矩阵, 用 1~9 标度法确定矩阵元素值, 计算各矩阵的相对权重, 结果见以下各表。

$P_{C_1-C}$	G 结果	$C_1$	$C_2$	$C_3$	$\alpha$
	$C_1$	1	1/3	2	0.230
	$C_2$	3	1	5	0.648
	$C_3$	1/2	1/5	1	0.122

$\lambda_{\max}=3.0037; C.R=0.0032<0.1$

$P_{C_1-I}$	$C_1$	$I_1$	$I_2$	$I_3$	$W_1$
	$I_1$	1	1/3	1/5	0.105
	$I_2$	3	1	1/3	0.258
	$I_3$	5	3	1	0.637

$\lambda_{\max}=3.0385; C.R=0.033<0.1$

$P_{C_2-I}$	$C_2$	$I_1$	$I_2$	$I_3$	$W_2$
	$I_1$	1	1/3	1/7	0.081
	$I_2$	3	1	1/5	0.188
	$I_3$	7	5	1	0.731

$\lambda_{\max}=3.065; C.R=0.056<0.1$

$P_{C_3-I}$	$C_3$	$I_1$	$I_2$	$I_3$	$W_3$
	$I_1$	1	2	3	0.592
	$I_2$	1/2	1	5	0.333
	$I_3$	1/7	1/5	1	0.075

$\lambda_{\max}=3.013; C.R=0.012<0.1$

由上各表可以进行组合权重计算:

	$C$	$C_1$	$C_2$	$C_3$	$V$
$I$	$\alpha$	0.230	0.122	0.648	
$I_1$		0.105	0.081	0.592	0.418
$I_2$		0.258	0.188	0.333	0.298
$I_3$		0.637	0.731	0.075	0.284

其中  $V1=0.230 \times 0.105 + 0.122 \times 0.081 + 0.648 \times 0.592 = 0.418$

3) 一致性检验。

检验结果显示 4 个判断矩阵均满足一致性要求。

4) 计算组合权重, 选择执行机构方案。

从组合权重的计算结果来看, 三个方案的优劣顺序为  $I_1, I_2, I_3$ , 所以应选用  $I_1$  圆柱凸轮间歇运动机构。

## 5 形态综合法

### 5.1 形态综合的基本概念

形态综合法, 是一种系统搜索的方法。它的思维方式属于“穷尽法”。形态综合法是系统地对多种因素可能的排列组合进行搜寻, 找出一切可能存在的方案, 以免丢失有潜力的方案。

这种方法是由美国的兹威基提出。形态综合法的要点是将机械系统分成若干部分, 对每个部分先寻求

其可能的解法, 然后对它们的各种组合一一加以考虑, 并得出各种可能的方案。

如果机械系统分成的部分数量较多, 而且每个部分又有很多的解法, 那么它的组合方案数量会过于巨大, 造成“方案爆炸”。为了便于评价、决策, 需要限制每个部分解法数量, 使方案数不要太多。为便于选择综合最优方案, 一般应按好坏程度进行方案排序, 从较好的若干方案中加以选择。

形态综合法又称形态学矩阵法, 它将机械系统的各部分与其解法以矩阵形式列出, 并用组合方法获得各种解决方案。

图 47.2-6 表示了形态综合法的求解过程。即将机械系统分解成若干子系统, 通过寻求方法获得各子系统的解。再用形态学矩阵, 组合成若干个机械系统的方案。最后通过评价决策, 确定最佳机械系统方案。

### 5.2 子系统的求解

以机械运动系统为例, 它的子系统就是各个执行机构。执行机构的功用是完成某一执行动作。因此, 我们可按执行动作的类型将机构进行分类, 并按运动和动力特性要求对各类机构进行表述, 以便选择合适的执行机构。表 47.2-5 说明执行机构的特性和类别。

表 47.2-5 中具体的执行机构及其特性可以参见机械工业出版社的《现代机械设备设计手册》第 8 篇机构及其系统设计。从某种意义上说表 47.2-5 就是执行机构的解法目录的组成形式。每类执行机构中有不少具体的机构。

同时, 我们还可以将表 47.2-5 形式的内容建立知识库, 便于进行计算机辅助机械运动系统方案设计。

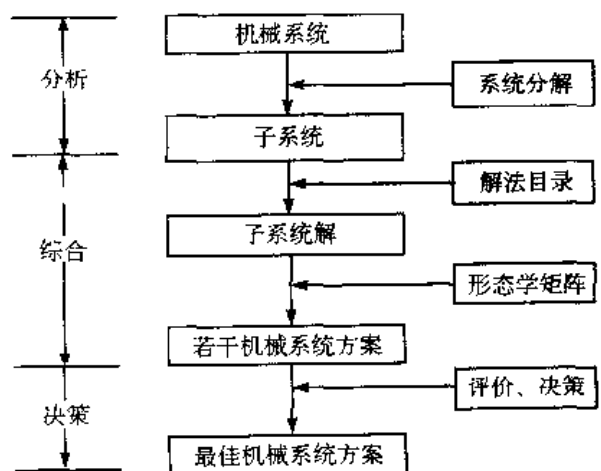


图 47.2-12 形态综合步骤

表 47.2-5 执行机构的特性和分类

机构类别	匀速转动机构	非匀速转动机构	往复移动机构	往复转动机构	间歇转动机构	间歇摆动机构	实现运动轨迹机构	刚体导引机构	实现其他功用机构
特 性	.....	.....	.....	.....	.....	.....	.....	.....	.....
运动类型及机构									
工作性能									
动力性能									
经济性									
结构紧凑性									

$$N = n_1 \times n_2 \times \dots \times n_i \times \dots \times n_m$$

5.3 形态组合法进行子系统解的组合

对于机械运动系统，我们将要求实现的执行动作列为纵坐标，将各执行动作的解列为横坐标，构成形态学矩阵如表 47.2-6 所示。

表中  $m$  为执行动作数目； $n_i$  为第  $i$  个执行动作可采用的执行机构个数。

根据组合原理，它能组出  $N$  种机械运动方案：

从上述计算式可求出可能采用方案数目往往会较大，说明设计人员必须从大量方案中选择和确定某一综合最优的方案。

例 2-3 确定行走式挖掘机的原理方案。

行走式挖掘机的总功能是取运物料，为了实现此总功能应由五个功能元来实现。因此行走式挖掘机的形态学矩阵列于表 47.2-7。

表 47.2-6 机械运动系统解的形态学矩阵

执行机构	1	2	3	4	...	$n_i$
执行动作						
$V_1$	$M_{11}$	$M_{12}$	$M_{13}$	$M_{14}$	...	$M_{1n}$
$V_2$	$M_{21}$	$M_{22}$	$M_{23}$	$M_{24}$	...	$M_{2n}$
⋮				⋮		
$V_i$	$M_{i1}$	$M_{i2}$	$M_{i3}$	$M_{i4}$	...	$M_{in}$
⋮				⋮		
$V_m$	$M_{m1}$	$M_{m2}$	$M_{m3}$	$M_{m4}$	...	$M_{mn}$

表 47.2-7 行走式挖掘机的形态学矩阵

功能元 (子系统)	功能元的解					
	1	2	3	4	5	6
A 动力机	电动机	汽油机	柴油机	蒸汽透平	液压马达	气动马达
B 移位传动	齿轮传动	滑轮传动	带传动	链传动	液力耦合器	
C 移位方式	轨道和车轮	轮胎	履带	气垫		
D 取物传动	拉杆	绳传动	气缸传动	液压缸传动		
E 取物方式	挖斗	抓斗				

由此可求出可能组合的方案数：

$$N = 6 \times 5 \times 4 \times 4 \times 2 = 960$$

根据多方面分析，可以采用以下二种方案：

$$A_1 + B_3 + C_3 + D_2 + E_1 \Rightarrow \text{履带式挖掘机}$$

$$A_5 + B_5 + C_2 + D_1 + E_2 \Rightarrow \text{液压轮胎式挖掘机}$$

例 2-4 确定三面自动切书机的原理方案

三面自动切数机是将装订好的书本的上、下及一侧切齐。因此它的执行动作有：送料动作—压书动作—上下切书—侧切书等四个执行动作。它的形态学矩阵见表 47.2-8。

表 47.2-8 三面自动切书机的形态学矩阵

执行动作	可能选择的执行机构			
	1	2	3	4
A 送料动作	凸轮机构	曲柄滑块机构	齿轮-齿条机构	多杆往复移动机构
B 压书动作	凸轮机构	肘杆机构	螺旋机构	
C 上下切书动作	凸轮机构	平面四连杆机构	平面多杆机构	
D 一侧切书动作	凸轮机构	平面四连杆机构	平面多杆机构	



因此,三面自动切书机的方案数为:

$$N=4 \times 3 \times 3 \times 3=108$$

#### 5.4 求最佳系统方案

在许多方案中进行比较,一般均应由粗到细,由定性到定量进行选优。首先进行粗筛选,把与设计要求不

符的或各功能元解不相容的方案除去。例如行动式挖掘机,若选用电动机。显然与液力耦合器、气垫、液压缸传动等不相容,无法组成可实现的方案。

在定性选取比较满意的几个方案后再用科学评价方法进行定量评价,从中选出符合设计要求的最佳方案(评价方法见第7章)。

## 第3章 动作行为载体及其创新设计

### 1 机械系统的功能—行为—结构特点

机械系统在总功能分解之后,对于分功能的求解目前常采用功能—行为—结构的求解步骤,即由功能求解实现功能的行为,由行为来构思实现行为的具体结构。

#### 1.1 总功能与工艺动作过程

对于机械运动系统来说,它的总功能是由工艺动作过程来实现的。工艺动作过程实际上是体现工作原理、工作过程和工作特点,是对机械系统总功能的较为具体的描述。因此,工艺动作过程的拟定是机械运动系统设计的关键。

机械运动系统的总功能是完成核心功能所需的一系列功能的总和,由总功能来确定相应的工艺动作过程的方法主要有:

##### (1) 基于实例的完善和改进

为了实现总功能而确定相应的工艺动作过程,可以首先选定与总功能相似的实例来进行分析,以确定工艺动作的程序。例如啤酒瓶灌装的总功能可找到相似的实例——汽水瓶灌装机,它的工艺动作过程的顺序为

瓶  
瓶盖 } 的储存与输送 → 汽水灌入瓶中 → 加盖及封  
汽水 }

口 → 贴商标 → 瓶装汽水的输出。

两者差别只在于瓶的外形及容量、商标的大小等,两者只是有较少的差异,有利于构思啤酒瓶灌装工艺动作过程。

又例,为构思香皂包装总功能,可以找到书籍包装机的工艺动作过程,其顺序为

书  
包装纸 } 的储存与输送 → 将包装纸裹包在书上的

几个动作 → 包装纸的粘贴 → 加贴外面的标签 → 包装后书的输出。

香皂包装与书籍包装两者的差别除大小不同外,还有外形的不同,因此需要修改其中包装纸裹包的几个动作和包装纸的粘贴动作。

总之,基于实例的完善和改进属于“举一反三”、“触类旁通”。这与一个人的知识和经验的积累有关。

##### (2) 拟人动作的分析

工作机器的工艺动作过程不少是模拟人的工作过程来构思的。例如,平版印刷机实际上是模拟人在纸上盖图章,因此就有:上墨—移动铅字版—印刷—取出印好纸张等。只要将这一动作过程适当加以完善就可作为平版印刷机的工艺动作过程。

拟人化、仿生化可以帮助我们构思工艺动作过程,这就是自然界的启示。

##### (3) 分功能动作求解的综合

机械运动系统总功能分解可得一系列分功能,采用相应的动作来综合所得的一系列动作过程,就可构成机械运动系统的工艺动作过程。不同的总功能分解方式就可以得到不同的工艺动作过程。通过分析比较就可求出更为适合的工艺动作过程。

分功能的动作求解,与分功能的工作原理密切相关。例如,“螺栓加工”可采用“车削”或“搓螺纹”。对于车削的分功能可分为:送料—车螺纹—切割—下料;对于“搓螺纹”的分功能可分为:送料—搓螺纹—下料。前者是将棒料送进,后者是将半成品送入搓螺纹工位。车螺纹的动作是车削,搓螺纹的动作是来回搓动。

#### 1.2 行为与执行动作

行为是功能的具体描述,行为本身就有广泛的含义。在机械运动系统中行为的具体表现就是动作。讲得更明确一点就是机器的执行动作。

工艺动作过程的分解与总功能的分解在机械运动系统中往往是一一对应的。因此,每一分功能就对应一个行为。对机械运动系统来说一般是对应一个执行动作。

在机械运动系统中能产生的执行动作种类是比较有限的,一般有:等速转动、不等速转动、往复摆动、往复移动、间歇转动、间歇移动、平面复杂运动(刚体导引)、空间复杂运动(空间刚体导引)、实现轨迹运动……。

工作就是行为,使我们更加理解机械运动系统设计的特点,明确设计的目标。

动作就是行为,也就要求我们在构思实现总功能的工艺动作过程时,应考虑每一动作实现的可能性。

#### 1.3 结构与执行机构

在功能—行为—结构的过程模型中,结构是功能

的载体、是行为的具体发生器。对于机械运动系统来说,结构不是别的,就是形形色色的机构,是产生执行动作的机构。或者就称之为执行机构。产生执行动作的执行机构,光是传统的刚性机构就不下千种。这在各种各样的机构手册中就可以见到。这些机构大多在现有的机器中广泛应用。同样,在今后创新机器中也可以采用。

现有的机构是选择执行机构时可靠的功能载体。创新从未有过的新机构也是寻求执行机构的重要途径。

随着现代机构概念的产生,执行机构已不仅仅限于传统的刚性构件机构,还有考虑弹性构件、挠性构件的机构。同时,还有各种各样单自由度、多自由度的可控机构。

因此,可以这样说,执行机构的不断创新是机器创新的基础,是功能求解的取之不尽用之不竭的源泉。

#### 1.4 工艺动作过程—执行动作—执行机构的功能求解模型

由通用性较强的功能—行为—结构(FBS)功能求解模型发展到针对性较强的工艺动作过程—执行动作—执行机构(PAM)功能求解模型,使机械运动系统设计与机构学紧密结合起来。使机构学从原来重点研究单个机构转向同时研究机构系统的问题。同时还使机构学与现代机械设计方法学结合在一起。这无疑是一种创新,推动了机构学的发展。

由于PAM功能求解模型具有较强针对性;执行动作与执行机构又有一定规律映射性;各执行机构又可进行一定程度上的可比性。因此,这一功能求解模型将会有利于开展计算机辅助设计,使机械系统设计有可能进入一定程度的智能化、自动化。从这个意义上看PAM功能求解模型可以推动机械系统设计创新。

## 2 动作行为和执行机构

对于机械产品中机械运动系统,其功能分解过程也就是根据工作原理来构思工艺动作过程。人们设计新机器是为了完成某种生产任务,机械运动系统的设计目的是要实现这种工艺动作过程。整个工艺动作过程往往可分解为若干个动作行为或运动行为,工艺动作过程就是由这些动作行为按一定顺序来完成。

### 2.1 常见的动作行为形式

机械运动系统常见的动作行为有旋转运动、直线运动、曲线运动以及空间曲线和刚体导引运动等。

#### (1) 旋转运动

1) 连续旋转运动 如车床、铣床的主轴以及缝纫机的上轴的转动等;

2) 间歇旋转运动 如自动车床工作台的转位,步进滚轮的步进运动等;

(3) 往复摆动 如颚式碎矿机的动颚板的运动,电风扇的摇头运动等。

#### (2) 直线运动

1) 往复移动 如压缩机的活塞、冲床的冲头的运动等;

2) 间歇往复移动 如自动机床或半自动机床的刀架运动等;

3) 单向间歇直线移动 如刨床的工作台进给运动等。

#### (3) 曲线运动

一般指执行构件上某点作特定的曲线运动。如缝纫机的送布牙作近似矩形轨迹运动,插秧机秧爪作近似人手插秧的曲线动作,电影放映机的抓爪作局部直线运动等。

#### (4) 刚体导引运动

一般指非连架杆的执行构件的刚体导引运动。如造型机工作台的翻转运动,折叠椅座位的导引运动等。

除上述执行构件的运动形式外,还有其他特殊功能的运动形式如微动、补偿、换向等。

遵循机器的运动方案设计要求,不同动作行为形式就应选择不同的合适的载体,也就是要找出能完成规定运动要求的物理(技术)装置。随着学科交叉的不断加剧和新的科学原理的应用,能完成某一动作行为的载体形式不断增多。动作行为载体呈现出门类众多、不断创新的特点。

### 2.2 动作行为载体(执行机构)的类型

动作行为载体总是表现为某种物理(技术)装置,在机构学中就是指各种各样的机构。该技术装置或机构中具体实现该动作行为的构件,传统上称为执行构件。在科学技术不断发展的形势下,机构已发展成为包含机、电、液、气、磁等多种实现原理,以实现机械运动为目的的广义机构。因此,在现代机构学中,动作行为载体就是广义机构。广义机构是在现代条件下对传统机构的拓展和延伸。广义机构的运动输出构件也就是执行构件。

运动行为载体即广义机构的分类有多种方式,可按结构形式、工作原理、机构功能等进行分类。传统的机械原理教科书上多是按结构形式进行分类,包括连杆机构、齿轮机构、凸轮机构、棘轮机构、槽轮机构、

摩擦轮机构、离合器机构、螺旋机构、组合机构、瞬心线机构、万向联轴节、共轭曲线机构等。按工作原理来分,则可分为传统机构(纯机械机构)、机液结合机构、电磁机构、机电结合机构、气动机构、光电机构、声电机构、记忆合金机构、机电液结合机构、电气机构等,其中除传统机构外的其他机构组成意义上的广义机构,相当一部分为电子、液、气、声、光、电、磁材料等学科领域技术在传统机构中的应用结果。鉴于机械运动系统的设计多从功能入手,为便于设计人员工作,越来越多的设计手册以功能来对各类机构进行分类。机构的功能主要体现在执行构件的功能上,而执行构件的功能通过运动来实现。因此从本质上说机构的功能通过运动来实现,也就是机构的功能是运动功能,根据运动目的的不同,执行构件在机械运动方案中实现的功能主要分为两大类:

#### (1) 实现运动形式变换的机构

在绝大多数机械中,原动机的运动形式为转动,执行构件的运动形式多种多样,主要有:

1) 匀速与非匀速转动机构。其中匀速转动机构包括定传动比平行轴转动机构(圆柱齿轮机构、平行四边形机构、双转块机构、同步齿形带传动机构、周转轮系机构等)、定传动比相交轴转动机构(锥齿轮传动机构、双万向联轴器等)、定传动比交错轴转动机构(螺旋齿轮传动机构、准双曲面齿轮传动机构、蜗轮蜗杆传动机构等)、可调传动比匀速转动机构(定轴轮系、周转轮系等)等;非匀速转动机构包括双曲柄机构、转动导杆机构、单万向联轴器、齿轮连杆机构、非圆齿轮机构等。

2) 往复移动和往复摆动机构。往复移动机构包括曲柄滑块机构、齿轮齿条机构、移动从动件圆柱(锥)凸轮机构、移动从动件盘形凸轮机构、直线电机等;往复摆动机构包括摆动从动件圆柱(锥)凸轮机构、摆动从动件盘形凸轮机构、曲柄摇杆机构、摆动导杆机构、伺服电机等。

3) 间歇运动机构。间歇运动机构包括棘轮棘爪机构、槽轮机构、凸轮分度机构、带停歇段的凸轮机构、不完全齿轮机构、伺服电机等。

4) 实现预期运动轨迹的机构。包括平面铰链四杆机构、组合机构、机电组合机构等。

5) 刚体导引机构。

6) 运动复合机构。

#### (2) 实现其他功能的机构

在机械中为了调整、操纵机器方便,改善机器的运转质量,保障机器的安全等目的,主要有:

1) 实现运动方向变换功能的机构。包括换向机构、单向机构、超越机构、双向电机等。

2) 实现运动离合或开停功能的机构。

3) 实现过载保护功能的机构。包括摩擦轮机构、带传动机构、止动开关等。

4) 实现微动和补偿的机构。包括螺旋差动机构、谐波传动机构、差动轮系机构、杠杆式差动机构、圆柱凸轮蜗杆误差补偿机构、轴线位置误差补偿机构、伺服电机驱动机构等。

5) 增力机构。

很显然,当知道某一机器的工艺动作过程及其工作要求之后,就可以根据所需实现某一动作与运动要求来选择合适的机构。在机电一体化等新技术应用不断广泛的今天,上面列出的各种功能实现的物质载体已不仅仅局限于纯机构,更多地体现为机、电、液、气、光、磁、声等多学科技术融合而产生的新型机构,即广义机构了。

选择合适的机构可参见由孟宪源主编的《现代机构手册》及辛一行、邹慧君等主编的《现代机械设备设计手册》第1卷第8篇“机构及其系统设计”。

## 3 机构组合和组合机构

在生产实际中,对机构的运动形式,运动规律及动力性能等有各式各样的要求,常见的齿轮机构、凸轮机构和连杆机构等基本机构往往不能满足一些场合的应用,于是要把一些基本机构组合起来,或者干脆把一些基本机构组合成一种与原基本机构特点不同的新的复合机构,即组合机构来实现生产中某些特殊的工艺要求。随着生产的不断发展,机构组合的形式和组合机构的种类将越来越多,更多的新机构有待发掘和创造。

机构组合的方式很多,常见的有:串联式、并联式、叠合式、叠联式等。

组合机构,常见的有:齿轮-连杆机构、凸轮-连杆机构、凸轮-齿轮机构等。

### 3.1 机构的串连式组合

把2个或2个以上的基本机构通过串连方式连接起来称为串连式组合。它可分为2种形式。

#### (1) 构件固接式串连

如果把若干个1个自由度的基本机构,将前一个机构的输出构件和后一个机构的输入构件固接(即前一个机构的输出构件就是后一个机构的输入构件),这种组合方式就称为构件固接式串连。如图47.3-1所示为钢锭热锯机构。它由双曲柄机构和曲柄滑块机构串连而成,双曲柄机构的输出曲柄就是曲柄滑块机构的输入曲柄。该机构能满足锯条(即滑块)在工作行程时作近似等速运动,而回程时具有急回特性的要求。它的

结构简单，急回系数大，生产效率高。又如图 47.3-2 所示辊筒式平版印刷机的自动送纸机构。它采用一对椭圆齿轮机构与曲柄滑块机构串连而成。从动的椭圆齿轮 2 与曲柄滑块机构中的曲柄固接，它既是椭圆齿轮的输出构件又是曲柄滑块机构的输入构件。当纸张送入印刷辊筒前，要求送进速度最慢以便纸张校准，防止它被压皱；而当纸张进入辊筒时，则要求它的速度与辊筒的圆周速度尽可能一致，以保持同步。采用椭圆齿轮和曲柄滑块机构串连而成的组合方式便能满足送纸工艺要求。

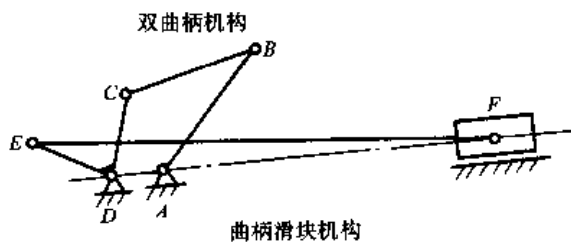


图 47.3-1 钢锭热锯机构

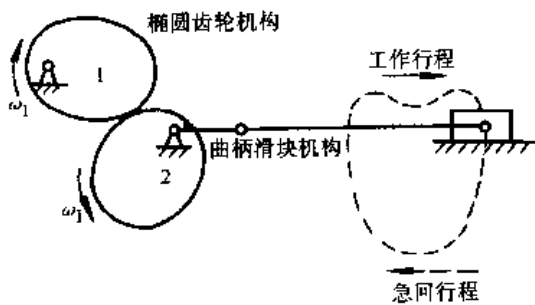


图 47.3-2 印刷机的自动送纸机构

(2) 轨迹点串连

假若前一个基本机构的输出为平面运动构件上某一点  $M$  的轨迹，通过轨迹点  $M$  与后一个机构相连，这种连接方式称为“轨迹点串连”。如图 47.3-3 所示织布机开口机构。它由曲柄滑块机构和转动导杆机构通过  $M$  点和滑块 4 用铰链串连而成。当曲柄以等角速转动时，连杆上的  $M$  点画出一条连杆曲线，而当转动导杆机构的输入构件滑块 1 上的  $M$  点沿此连杆曲线的直线段  $EF$  上运动时，构件 2 就能实现较大停歇的运动要求。

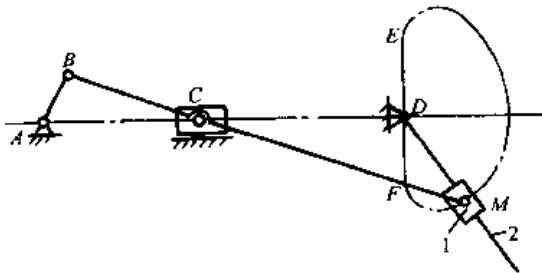


图 47.3-3 织布机的开口机构

### 3.2 机构的并联式组合

并联式组合的机构，不少属于复合式的。

下面介绍 2 种并联式组合机构的形式。

1) 把原动件的一个运动同时输入给  $n$  个并列布置的单自由度机构，再转换成  $n$  个输出运动；这  $n$  个运动又输入给同一个  $n$  自由度的基本机构，然后再合成一个输出运动，这种组合方式称并联式组合。如图 47.3-4 所示的铁板输送机构，它由定轴齿轮机构和曲柄摇杆机构通过差动轮系机构并联组合，使定轴齿轮机构的输出运动  $\omega_1$  和曲柄摇杆机构的输出运动  $\omega_2$  作为差动轮系机构的输入运动，然后合成为一个输出运动  $\omega_3$ ，它可以使与轮 7 固联的送料辊获得具有短暂停歇的送进运动。

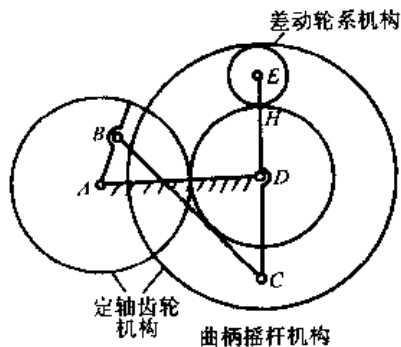


图 47.3-4 铁板输送机构

2) 把 2 个或 2 个以上单自由度的基本机构共用一个输出构件输出运动，这也是一种并联式组合。如图 47.3-5 所示的星形发动机机构，它由 6 个曲柄滑块机构组成。6 个活塞的往复运动同时通过连杆传给公用曲柄  $AB$ ，其输出转动是 6 个曲柄滑块机构输出转动的代数和。与单缸发动机相比，它的输出扭矩波动小，可以部分地或全部地消除振动力。

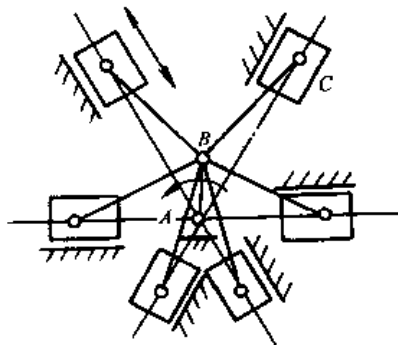


图 47.3-5 6 缸星形发动机机构

### 3.3 机构的叠合式(或运载式)组合

把一个机构叠装在另一个机构的构件上，两机构

各自进行运动,其输出运动则由两机构运动叠加而成。这种组合方式称为机构的叠合式(或运载式)组合。如图 47.3-6 所示的电动玩具马的主体运动机构。它能模仿马的奔驰运动形态,使玩具马上的小朋友仿佛身临其境。实际上,这种电动马由曲柄摇杆机构叠加在两杆机构绕  $O-O$  轴转动的构件 4 上。两杆机构在此作为运载机构使马绕以  $O-O$  轴为圆心的圆周向前奔驰;而曲柄摇杆机构中的导杆 2 的摇摆和伸缩则使马获得跃上、窜下、前俯后仰的姿态。

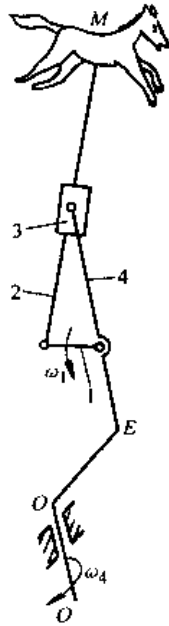


图 47.3-6 玩具马机构

### 3.4 机构的叠联式组合

把后一个基本机构叠联在前一个基本机构上,称为机构的叠联式组合。如图 47.3-7 所示,它是 1 台全液压的挖掘机,其挖掘动作由 3 个带液压缸的基本连杆机构组合而成。它们一个紧挨着一个,而且后一个基本机构的相对机架正好是前一个基本机构的输出构件。挖掘机臂架 3 的升降、铲斗柄 7 绕  $D$  轴的摆动以及铲斗 10 的摆动分别由 3 个液压缸驱动,它们分别或协调动作时,便可使挖掘机完成挖土、提升和倒土等动

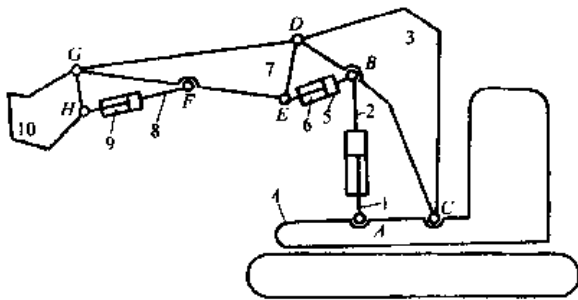


图 47.3-7 挖掘机的作业机构

作。

### 3.5 组合机构

组合机构是由凸轮构件、连杆构件和齿轮构件相互组成的复合机构,其分析和设计方法与单纯的凸轮机构、连杆机构和齿轮机构不同。常见的有凸轮连杆机构、齿轮连杆机构和凸轮齿轮连杆机构。

#### (1) 凸轮连杆机构

图 47.3-8 所示为凸轮连杆机构型的打字机,凸轮 1 为主动件,摆杆 4 (打字机) 为从动件。为了使打字杆能以一定的冲力撞击到压纸筒上,其角速度开始较小,将要撞击时较大,故要求从动件具有一定的加速度。

#### (2) 齿轮连杆机构

图 47.3-9 所示为机械手抓取机构,它是齿轮连杆机构,它由曲柄摇块机构 1-2-3-4 与齿轮 5、6 组合而成。齿轮机构的传动比等于 1。活塞杆 2 为主动件,当液压推动活塞时,驱动摇杆 3 绕  $A$  点摆动,齿轮 5 与摆杆 3 固定,并驱使齿轮 6 同步运动。机械手 7、8 分别与齿轮 5、6 固结,可以实现铸工搬运压铁时夹持和松开压铁的动作。

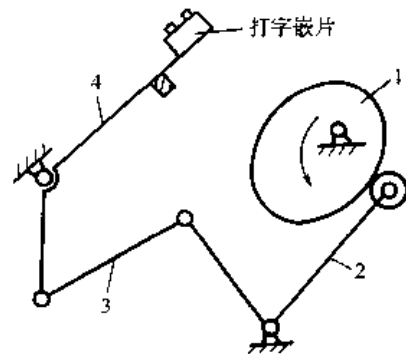


图 47.3-8 凸轮连杆型的打字机

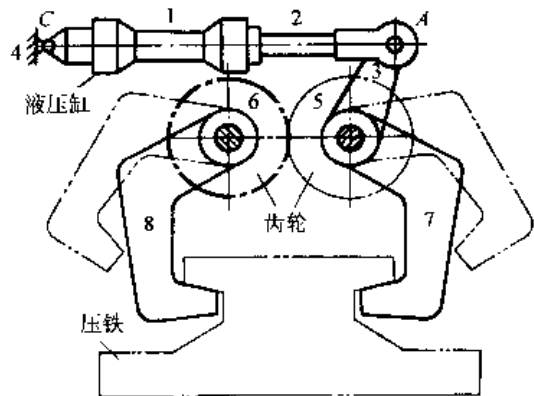


图 47.3-9 机械手抓取机构

#### (3) 凸轮齿轮连杆机构

图 47.3-10 所示为穿孔机构,它是凸轮齿轮连杆

机构。构件1、2为具有凸轮轮廓曲线并在廓线上制成轮齿的凸轮齿轮构件。构件1与手柄相固接。当操纵手柄时，依靠构件1和2凸轮廓线上轮齿相啮合的关系驱使连杆3、4分别绕D、A摆动，使E、F移近或移开，实现穿孔的动作。

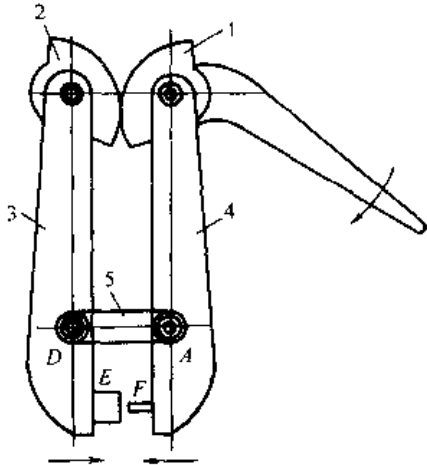


图 47.3-10 穿孔机构

#### 4 广义机构

科学技术的迅速发展使得机械的构成发生了很大的变化，引入液、气、声、光、电、磁等工作原理的新型机构应用日益广泛。将这类不同于传统机构的现代机构统称为广义机构。广义机构是一些在实现工作原理和结构型式等方面跨越了纯机械领域限制、有所创新的机构。由于利用了一些新的工作介质或工作原理，广义机构可比传统机构更简便地实现运动或动力转换。广义机构还可以实现传统机构较难以完成的运动。广义机构种类繁多，往往是由包括机械在内的多种学科原理的交叉融合而得来。按工作原理的不同广义机构可以分为液动机构、气动机构、电磁机构、振动机构、光电机构、声电机构、记忆合金机构、机电组合机构等。按结构型式及用途的不同广义机构又可以分为微位移机构、微型机构、信息机构、智能机构等。限于篇幅，下面将简要介绍应用较为广泛的部分广义机构的基本结构和特点。

##### 4.1 液动机构

液动机构是以具有压力的液体作为工作介质来实现能量传递与运动变换的机构，广泛应用于矿山、冶金、建筑、交通运输和轻工等行业。

###### (1) 液动机构的特点

液压传动与机械传动、气动传动等相比具有下述优点：

- 1) 易于无级调速，调速范围大；

- 2) 体积小、重量轻、输出功率大；

- 3) 工作平稳，易于实现快速启动、制动、换向等动作；

- 4) 控制方便；

- 5) 易于实现过载保护；

- 6) 由于液压元件自润滑、磨损小，工作寿命长；

- 7) 液压元件易于标准化、模块化、系列化。

液压传动也具有下述缺点：

- 1) 油液具有压缩性、易泄漏，易污染环境且传动不准确；

- 2) 液体对温度变化很敏感，不易在变温或低温环境下工作；

- 3) 效率较低，不易作远距离传动；

- 4) 制造精度要求高。

###### (2) 液动机构应用实例

图 47.3-11 为一机械手臂伸缩液动机构。它由数控装置发出指令脉冲，使步进电动机带动电位器触头转动一个角度  $\theta$ 。如果为顺时针转动，动触头偏离电位器中点，在其上的引出端便产生与指令信号成比例关系的微弱电压  $u_1$ ，经放大器放大为  $u_2$  作为信号电压输入电液伺服阀的控制线圈，使电液伺服阀产生一个与输入电流成正比例的开口量。这时压力油以一定的流量  $q$  经阀的开口进入液压缸左腔，推活塞连同机械手臂向右移动  $x$ 。液压缸右腔的油液经伺服阀流回油箱。由于电位器外壳上的齿轮与手臂上的齿条相啮合，因此手臂向右移动的同时，电位器逆时针方向转动。当电位器的中点与动触头重合时，动触头引出端无电压输出，放大器输出端的电压为零，电液伺服阀的控制线圈无电流通过，阀口关闭，手臂停止移动。反之，当指令脉冲的顺序相反，则步进电动机逆时针方向转动，手臂向左移动。手臂的运动速度决定于指令脉冲的频率，而其行程取决于指令脉冲的数量。

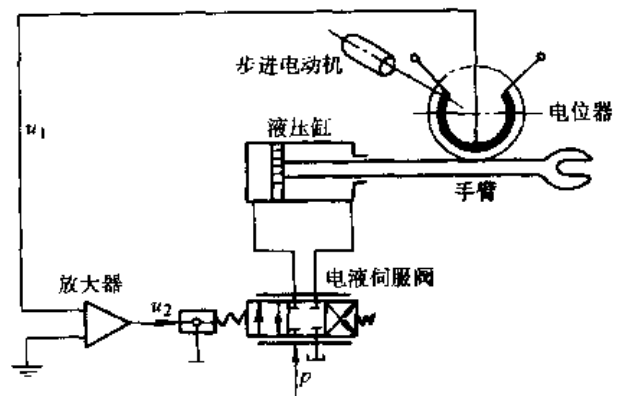


图 47.3-11 机械手臂伸缩液动机构

## 4.2 气动机构

与液动机构相类似,气动机构是以具有压力的气体作为工作介质来实现能量传递与运动变换的机构,同样广泛应用于矿山、冶金、建筑、交通运输和轻工等行业。

### (1) 气动机构的特点

气动机构具有下述优点:

- 1) 工作介质为空气,易于获取和排放,不污染环境;
- 2) 空气黏度小,故压力损失小,适于远距离输送和集中供气;
- 3) 比液压传动响应快,动作迅速;
- 4) 适于恶劣的工作环境下工作;
- 5) 易于实现过载保护;
- 6) 易于表征化、模块化、系列化。

### (2) 气动应用实例

图 47.3-12 所示为一种比较简单的可移动式气动通用机械手的结构示意图。由真空吸头 1、水平缸 2、垂直缸 3、齿轮齿条 4、回转缸 5 及小车等组成。可在 3 个坐标内工作。一般用于装卸轻质、薄片工件,只要更换适当的手指部件,还能完成其他工作。

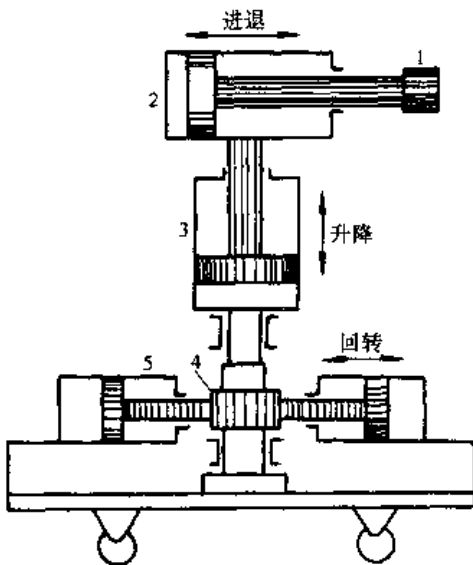


图 47.3-12 通用机械手结构示意图

## 4.3 电磁机构

电磁机构由电与磁的相互作用来完成所需的机械运动,拥有庞大的家族。最常见的电磁机构可以十分方便地实现回转运动、往复运动、振动等。它广泛应用于继电器机构、传动机构、仪器仪表机构、开关机构、电磁振动机构、电动按摩器、电动理发器、电动剃须刀中。

### (1) 电磁机构的特点

电磁机构具有下述优点:

- 1) 电磁致动,电源形式众多;
- 2) 驱动和执行件合二为一,结构简单;
- 3) 运动控制和调节方便。

### (2) 电磁机构应用实例

图 47.3-13 所示为电动锤机构示意图。利用 2 个线圈 1、2 的交替磁化,使锤头 3 产生往复直线运动。图 47.3-14 所示为电磁开关示意图。电磁铁 1 通电后吸合杆 2,接通电路 3。断电后,杆 2 在复位弹簧 4 的作用下,脱离电磁铁,电路断开。

另外还有反电磁机构,是利用机械运动的切割磁力线作用产生电信号,对电信号进行处理后可判断机械振动的位移大小和频率。反电磁机构多用于磁电式位移或速度传感器中。

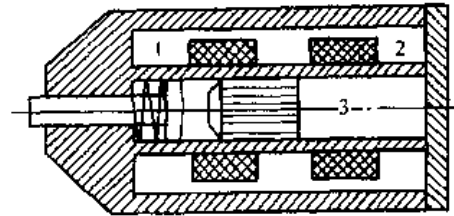


图 47.3-13 电动锤机构示意图

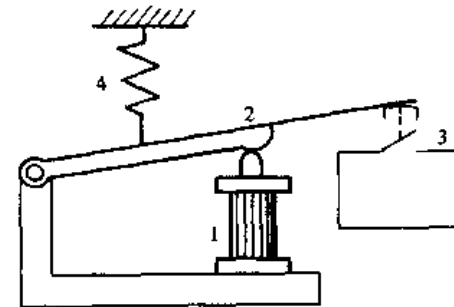


图 47.3-14 电磁开关机构示意图

## 4.4 振动机构

利用振动产生运动和动力的机构称为振动机构。用来产生振动的方式有电磁式、机械式、音叉式、超声波式等。振动机构在轻工业中获得广泛应用。对于各种小型产品(如钟表元件、无线电元件、小五金制品)、粉粒料(如味精、洗衣粉、食盐、糖等)、易碎物品(如玻璃、陶瓷制品等)等,振动机构都可以作为有效的送料机构。其中电磁振动机构应用最为广泛。

### (1) 振动机构的特点

电磁振动机构与其他送料机构相比,具有下述优点:

- 1) 结构简单、重量较轻;
- 2) 送料速度容易调节;
- 3) 物料移动平稳;



4) 消耗功率小;

5) 适用范围广。

(2) 振动机构应用实例

图 47.3-15 所示为圆盘型电磁振动送料机构示意图。该机构沿圆周装有 4 个电磁激振器, 每个电磁激振器均呈倾斜安装。用电磁激振器振力强迫漏斗 4 及底座 1 产生垂直运动和绕垂直轴的扭转振动。图中 2 为板簧, 3 为衔铁, 5 为线圈, 6 为铁心, 7 为橡胶减振器。振动频率通常为 3000 次/min, 机器在近共振状态下工作。图 47.3-16 所示为一利用机械振动来进行工作的插入式振捣器。它由带有增速齿轮的电动机 7、增速器 4、软轴 3 和偏心式振动棒所组成。电动机 7 通过增速器 4 和软轴 3, 将动力传递给偏心轴 2, 使振动棒 1 振动, 用来振捣混凝土。软轴 3 的另一个作用是使振动棒在任意位置进行工作。在电动机轴 5 和增速器大齿轮之间有防逆转作用的超越离合器 6。图 47.3-17 所示为一音叉式振动机构。当音叉 1 振动时, 它轮流地接通电磁铁 2 和 3。当电磁铁 2 激励时, 其两极把轮 4 的突出部 a 和 b 吸引过来, 致使轮 4 绕 A 回转某一角度; 这时突出部 c 和 d 接近电磁铁 3 的两极。如果现在接通电磁铁 3, 则它的两极吸引突出部 c 和 d, 轮子又在相同的方向回转。此外, 还有超声波振动机构(图 47.3-18)等。

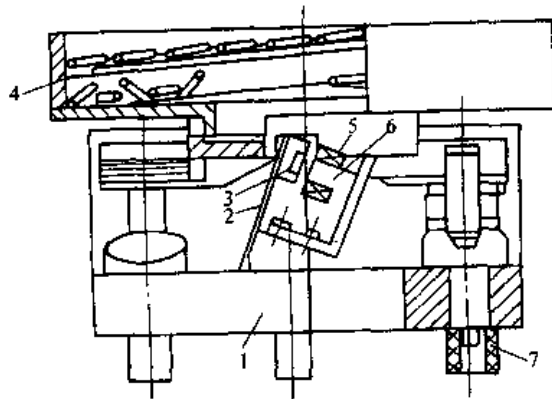


图 47.3-15 圆盘型电磁振动送料机构

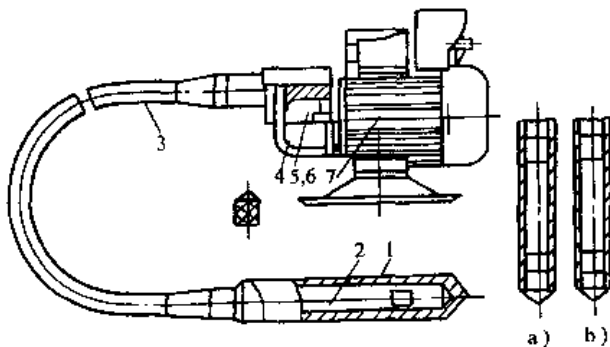


图 47.3-16 插入式振捣器

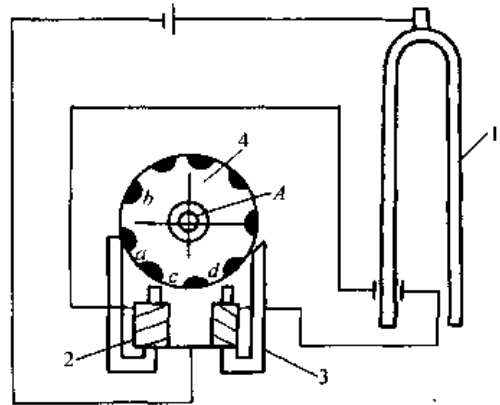


图 47.3-17 音叉式振动机构

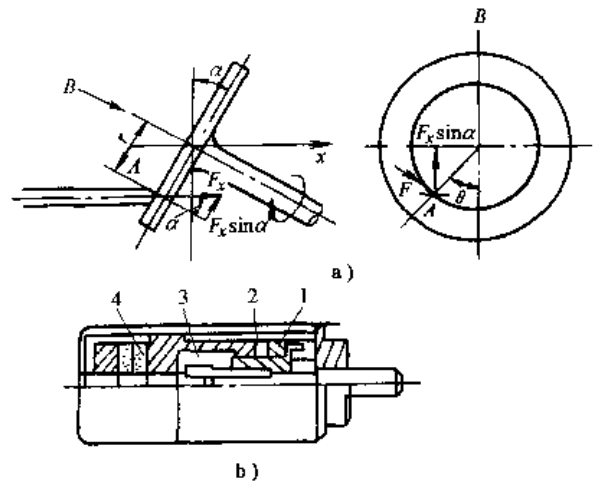


图 47.3-18 超声波振动机构

4.5 光电机构

(1) 光电机构的特点

光电机构具有下述优点:

- 1) 结构精巧、别致;
- 2) 广泛用作控制部件。

(2) 光电机构应用实例

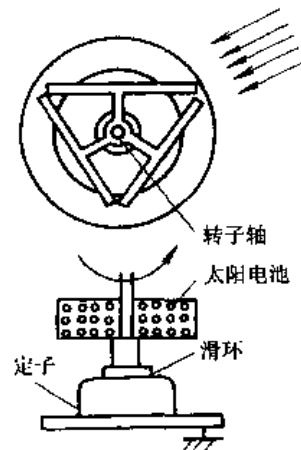


图 47.3-19 光电电动机原理图

图 47.3-19 所示为光电动机原理图，其受光面是一般太阳电池，使电动机的转子转动。图 47.3-20 是根据光化学原理将 NO<sub>2</sub> 分子数的变化转变为机械能的机构。

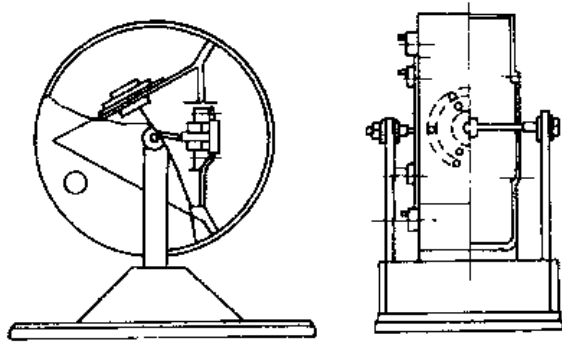


图 47.3-20 光化学回转活塞式星形马达

### 5 执行机构的创新方法

机械运动系统设计时，对系统方案构思与拟定过程中，选择执行机构的型式是非常重要的内容。通常情况可以考虑选用一些常用的、现有的机构型式。但是，

为了追求结构新颖、性能优良，应该采用机构创新设计方法进行执行机构的创新。

常用的机构创新设计方法有以下 7 种：

- 1) 应用机构学原理创新法；
- 2) 利用连杆或连架杆运动特点创新法；
- 3) 利用两构件相对运动关系创新法；
- 4) 利用成型固定构件实现复杂动作的创新法；
- 5) 利用多种驱动原理机构创新法；
- 6) 机构类型创新和变异创新法；
- 7) 机构类型替代创新法。

#### 5.1 应用机构学原理

##### (1) 叠加杆组创新机构

根据平面机构组成原理在一个机构上叠加一个或多个杆组后，便可以形成各种新的机构来满足运动转换或实现某种要求的功能。

如图 47.3-21 所示的发动机机构就是在曲柄滑块机构的基础上叠加 2 个 II 级杆组所组成。在四杆机构上叠加杆组不改变机构的自由度，却能增加机构的功能。例如，可取得有利的传动角、较大的机械利益，改变从动件的运动特性，增加从动件的行程等。

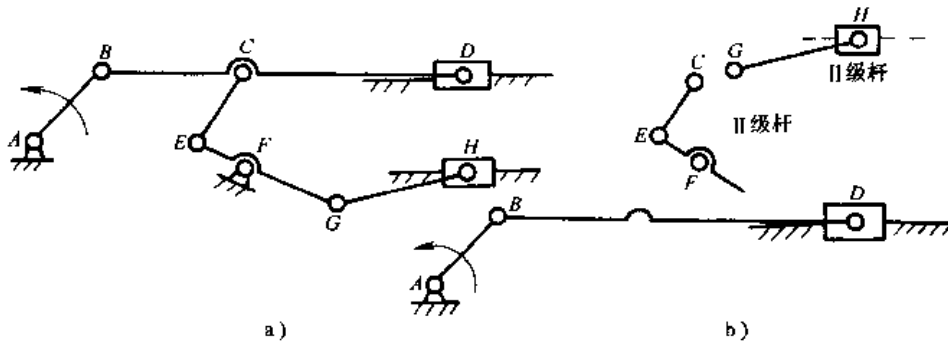


图 47.3-21 发动机机构

图 47.3-22 所示为一钢料推送机的机构运动简图，该机构是在铰链四杆机构 ABCD 上叠加了一个 II 级杆组所构成。变成多杆机构后，可使从动件 5 的行程大幅度扩大。

##### (2) 转动副扩大、高低副互代法

1) 转动副扩大 组成运动副的两元素可按同一比例任意扩大或缩小，而不影响该副的特性。图 47.3-23a 所示的机构中，偏心轮绕轴心 A 转动，它与构件左端具有同样直径的孔配合，其几何中心 B 在构件和上的位置不变，即这两个构件在 B 处组成转动副。当选择图 47.3-23b 所示的机构运动简图时，若因 AB 杆太短而无法安装两个转动副时，那么可在与其运动特性完全相同的图 47.3-23a 所示机构的基础上进行结构设

计，问题即可获得解决。

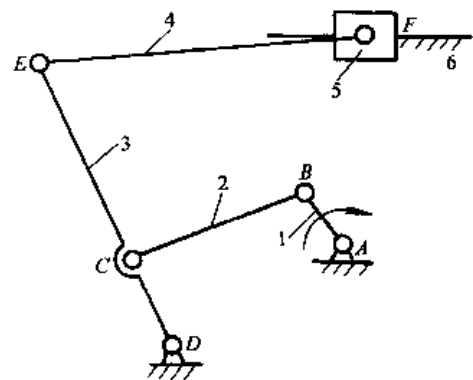


图 47.3-22 钢料推送机构

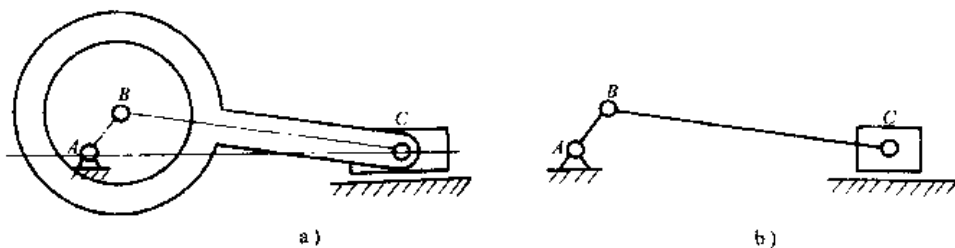


图 47.3.23 偏心轮机构

2) 高、低副互代 图 47.3-24 所示为高副低代。将高副代换后对其运动特性没有影响,且因低副是面接触,所以易于加工,可以提高耐磨性。

图 47.3-25 所示为低副高代。进行低副高代有利于实现复杂的输出运动规律。

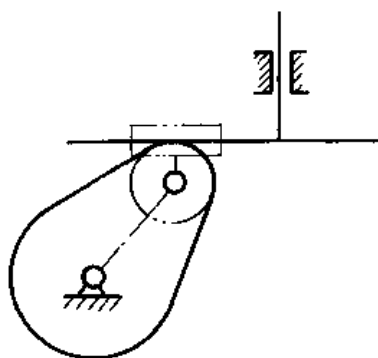


图 47.3-24 高副低代

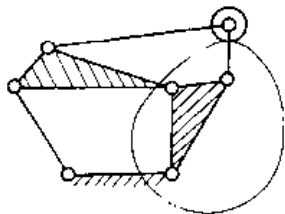


图 47.3-25 低副高代

(3) 其他现有原理的创新

包起帆发明的异步抓斗成功地解决大小不一的生铁块的抓住问题。由于其结构新颖、安全可靠,获奖颇多。

从异步抓斗的机构学原理来看,发明者创造性地应用了两项已有的成熟技术。但是这种抓斗的关键,第一是多自由度差动滑轮组;第二是复式滑块连杆机构。

多自由度差动滑轮组主要目的就是为了获得各个爪的差动运动,达到异步目的。

同理,可以将具有3个、4个或更多个自由度的差动滑轮,组合相应数目的复式滑块连杆机构,可组成3个、4个或更多个自由度和瓣数的异步抓斗机构。瓣瓣异步抓斗,采用沿圆周方向均布,这里不再赘述。这种花瓣式的异步抓斗可以抓取大小不一的铁块,通过瓣瓣的逐一闭合、并最后所有瓣瓣的完全闭合来做到

安全、高效的抓取物块。从实际效果来看,这种抓斗比原来同步抓斗要好得多。包起帆的发明再次说明,应用已有的原理同样可以创造出性能优良、动作巧妙的新颖机构。

5.2 利用连杆机构或连架杆机构的运动特点构思新的机构

利用简单机构的某些连架杆或连杆的运动特点完成某一动作过程是机构创新的一种有效方法。

图 47.3-26 所示的车门开闭机构为一反平行四边形机构,它利用反平行四边形机构运动时两曲柄转向相反的运动特点,使两扇车门同时打开或关闭。

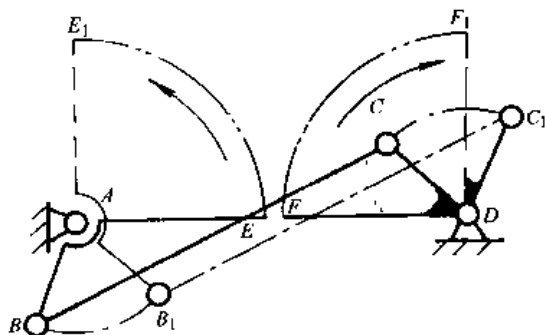


图 47.3-26 车门开闭机构

图 47.3-27 所示为铸造用大型造型机的翻箱机构。该机构应用双摇杆机构 ABCD,将固定在连杆 BC 上的沙箱位置进行造型震实后,转到 B'C' 位置,翻转

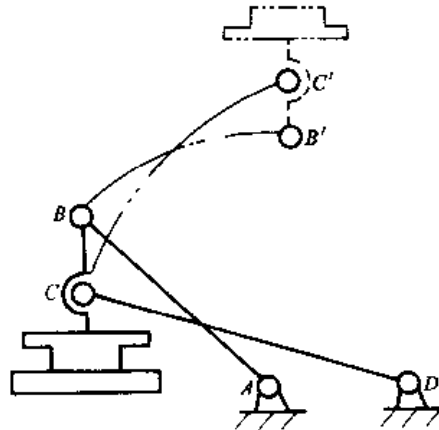


图 47.3-27 翻箱机构

180°，以便进行拔模。

图 47.3-28 所示的摄影平台升降机构为一平行四边形机构，它利用了连杆作平移运动的特性。图 47.3-29 所示为电影摄影机和放映机的拉片机构，为了使胶片获得间歇移动，采用了图示的曲柄摇杆机构。曲柄 1 转动时，摇杆 3 摆动，使作平面运动的连杆 2 上的点 E 走出近似 D 型的运动轨迹，拉片爪就能按图示箭头方向沿 D 型轨迹运动。使用时，拉片爪首先插入电影胶片 4 两侧的孔中，然后向下拉动胶片一段距离，接着爪退出齿轮，此时胶片静止不动，以待摄取或放映胶片上的画面，拉片爪再上行，重复上述运动循环，使胶片作间歇式的移动。图 47.3-30 所示为双缸压气机运动简图。单缸压气机为一曲柄滑块机构，当滑块(活塞)往复运动时，由加速度所产生的惯性力作用于机座 2 上，当曲柄 1 的转速很高时，惯性力产生的动载荷和振动都很大。为此，可把两组相同的曲柄滑块机构左右对称配置，使它们组成具有公共曲柄 1 的六杆机构，由于两滑块的加速度大小相同、方向相反，因而它们的惯性力对机座 2 的作用力可相互平衡。

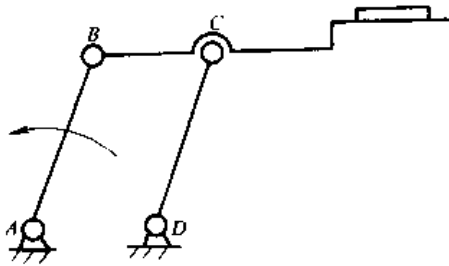


图 47.3-28 升降机构

总之，利用连架杆或连杆的运动特点来创新一些机构，对于一个设计人员来说还是十分重要的创新方法。

利用两构件相对运动关系来完成独特的动作过程，使机构创新有一种全新的思路。

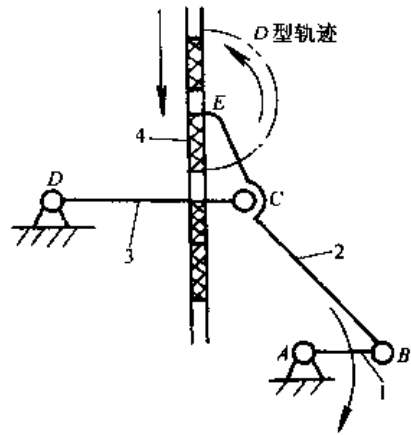


图 47.3-29 摄影机拉片机构

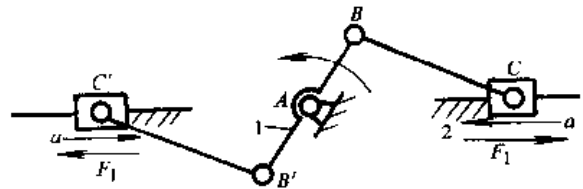


图 47.3-30 双缸压气机构

图 47.3-31 所示为一搬运物件机构。欲将一工件从输送道 A 上搬运到另一输送道 B 上，不但要使工件移位，其姿态还略有变化。按要求可以设想出很多搬运的方法，例如夹紧-平移-转位-松开。可利用工件上的孔将一根杆 3 沿孔轴线移动，将工件穿在杆 3 上(图 a)，然后将杆转动，使工件搬到 B 上(图 b)，最后将杆 3 退出(图 c)并转回原位(图 c)。这个搬运方法就只有先移动、再转动两个运动。图 47.3-31 所示机构是一种巧妙的结合，它在构件 5 上加一个摩擦块或滚珠定位器 6，主动杆为 1，推动挡块 2 沿取件杆 3 移动，杆 3 在顶起工件前杆 5 不动；杆 3 沿导轨移动，穿过工件孔并将其顶起，当杆 3 的挡块 2 碰到 5 的导轨端面在

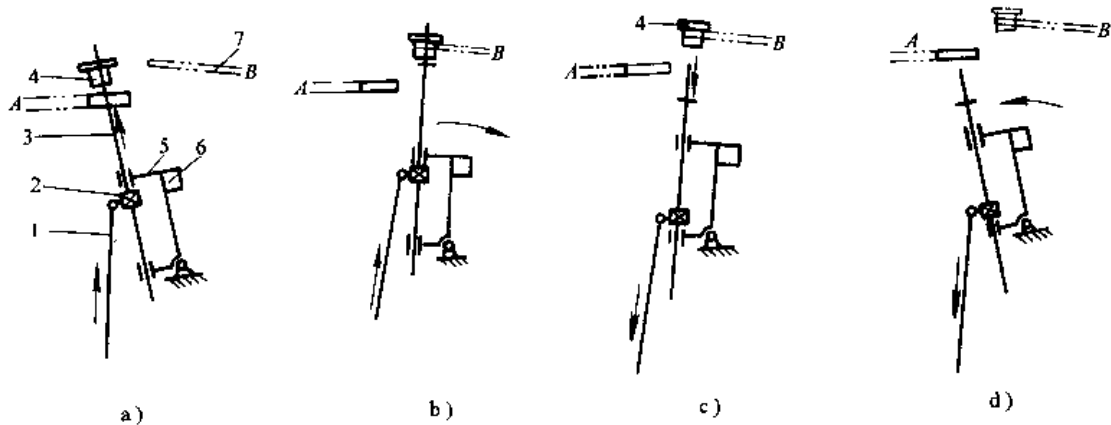


图 47.3-31 物体搬运机构

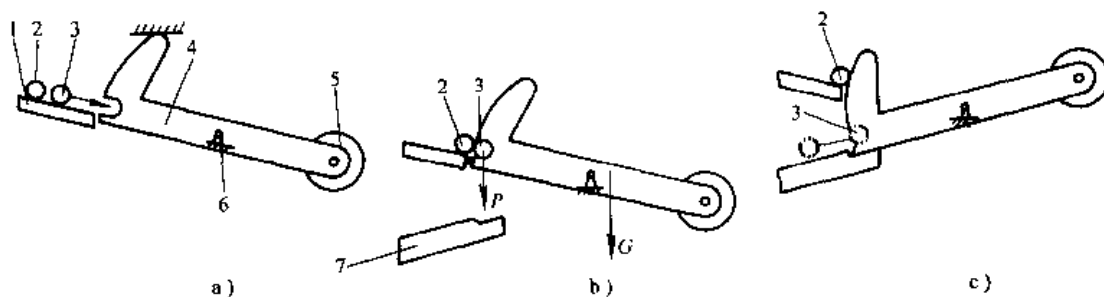


图 47.3-32 工件逐件转移机构

继续前行时,就推动杆5克服摩擦块或定位器6的阻力而转动,将工件送到B上,然后主动杆往回运动,杆3先沿工件孔的轴线移退出工件,再转至原位。图47.3-32所示为一个简单而有效的工件转移机构。上送料板1上的工件2及3被逐个移置于下送料板7上。机构没有专用的动力源,仅靠工件的重力位能进行工作。图47.3-32a至图47.3-32c表示工件转移过程,工件的重力使杆4摆动。工件离开杆4后,配重5使杆4复位。杆4摆动一次只送出一个工件。杆的摆动周期主要取决于工件重力 $P$ 对支点6的矩,杆4(包括配重5)的重力 $G$ 对支点6的矩和工件及杆对支点的惯性矩。

图47.3-33为涮羊肉用的羊肉切片机。目的是要将冻肉块切成均匀且极薄的肉片。如果仿照手工切法,则要有刀片往复切片的动作和刀片横向往复拉动的动作以及肉块间歇推进的动作。若采用图47.3-33的刨切形式,刨台面作往复移动,羊肉块靠重力压在台面上,兼起送进作用,则此机构的结构就可做得极其简单,而且容易实现刨切薄片要求。

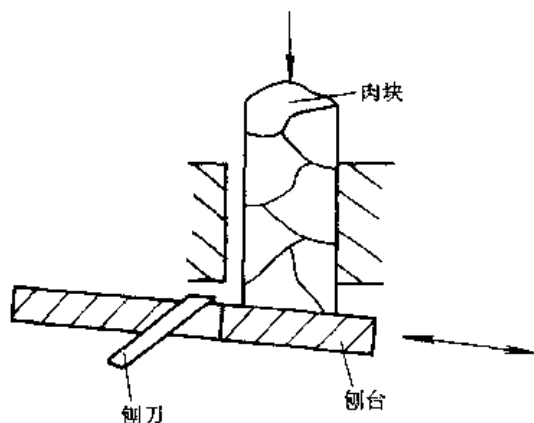


图 47.3-33 刨肉机

### 5.3 用成型固定构件和相对运动实现复杂运动过程

在轻工业生产过程中,如糖果、饼干、香烟、香皂等的裹包和颗粒状、液体状食品的制袋充填等,其工艺动作都比较复杂。为实现包装机械、食品机械等比较复

杂的工艺过程,如果按通常的工艺动作过程分解方法。对每个动作采用一个执行机构来完成,那么机械中的机构型式就很多,结构便很复杂。所以要求机构型式简单、合理、新颖,采用一些特殊形状固定模板来完成某些工艺动作,是方法之一。

图47.3-34所示为成卷塑料包装材料的供给机构。图中所示的装置中除成卷包装材料(支承在轴上)1、导辊组5、牵引输送辊4和裁切装置(图中未示出)外,还有检测装置2、成型器3(为一固定模板)和输送长度调节控制装置(图中未示出)。这种装置的特点是:由输送辊4牵引,由检测装置2通过识别包装材料上的标记,发出信号来控制进给输送辊及裁切装置。这样就可以保证每一次包装都可使包装材料上的图文处于相同的位置。

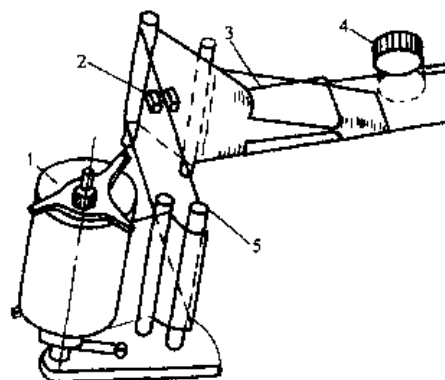


图 47.3-34 成卷塑料包装材料的供给机构

用成型固定构件来完成较为复杂的动作过程是一种有效的机构创新的方法。

### 5.4 利用多种驱动原理创新机构

在不少场合为了使机构新颖、独特、高效,往往要脱离纯机械模式而采用光、电、液等工作原理来创新机构。

图47.3-35所示为被检产品的电磁回转机构。当电磁铁1绕定轴D转动时,在被检测的金属零件2中感应出涡流,它和电磁场相互作用,电磁铁1产生转动

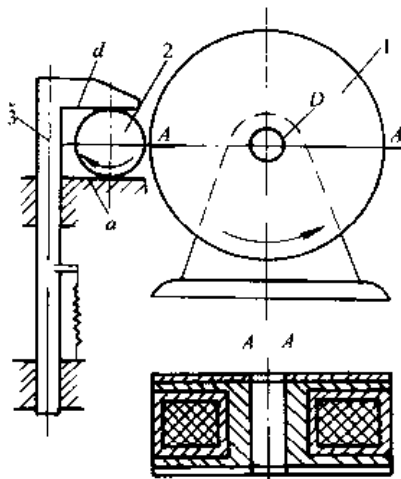


图 47.3.35 电磁回转机构

力矩，在电磁铁 1 转动的反方向转动被测工件。被测工件支承在静止的钳牙 *a* 上并且用可动的钳牙 *d* 压住，*d* 和测量器的心轴 3 固连。所以，这一回转机构可用于进行工件圆柱度检查的测量仪中。图 47.3-20 所示为光化学回转活塞式星型马达，它是一种根据光化学原理将  $O_2$  分子数的变化转变为机械能的特殊机构，用丙烯酸树脂制成的圆筒形容器内外周，被分隔成 3 部分作为反应室 1，室内装  $O_2$ 。反应室 1 的内侧壁上装有一曲柄滑块机构。介质受光照射后，由于光化学作用， $O_2$  的浓度发生变化而引起反应室压力变化，使活塞 2 运动并带动曲柄 3。相对于来自一个方向的太阳光，转动的各反应室自动地经过照射和背阴的反复循环，使曲柄做连续转动。如图 47.3-13 所示为电锤机构。直流电的电锤有一快速电流换向器，且每分钟冲击次数用电压进行调节。交流电的电锤每分钟有恒定的冲击数，它由所供电流的频率来决定。图 47.3-36 所示为液体的轴角速度均衡器。当与轴 *A* 固结的局部装有水银的筒管 1 转动时，轴 *A* 的角速度愈大，水银的质心离开轴线 *x-x* 愈远，此时，轴-筒系统的惯性矩增加。因此，依靠固有惯性矩的变化，使遭受周期性激振的轴的角速度均衡化。图 47.3-37 所示为液体用杠杆。盛液体

的容器 1(可绕轴 *A* 自由回转)上设置挡块 *a*，使其保持在作业位置上，当液体填满时，容器 1 下降，柱销 2 抵到挡板 3，容器翻转并倾泄出液体，倾泄出的液体质量由配重 4 决定。由图 47.3-38 可见，若只要求实现简单的工作位置变换，利用气缸作原动件就很方便。以图 47.3-38a 为例，要求摇杆实现 I、II 两个工作位置的变换，如利用曲柄摇杆机构，往往要用电动机带动一套减速装置驱动曲柄，为了使曲柄能停在要求的位置，还要有制动装置。如果改用气缸驱动，则结构将大为简化(如图 47.3-38b)。

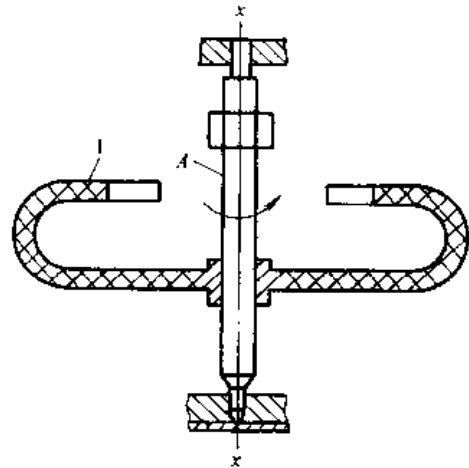


图 47.3.36 轴的角速度均衡器

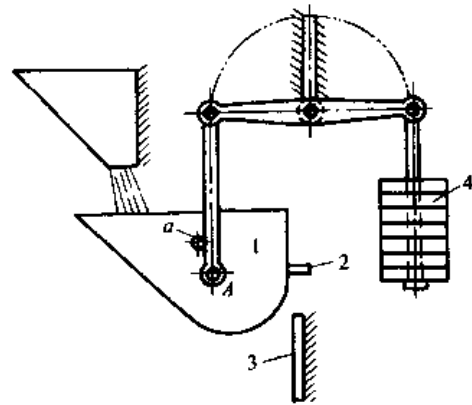


图 47.3.37 液体用杠杆

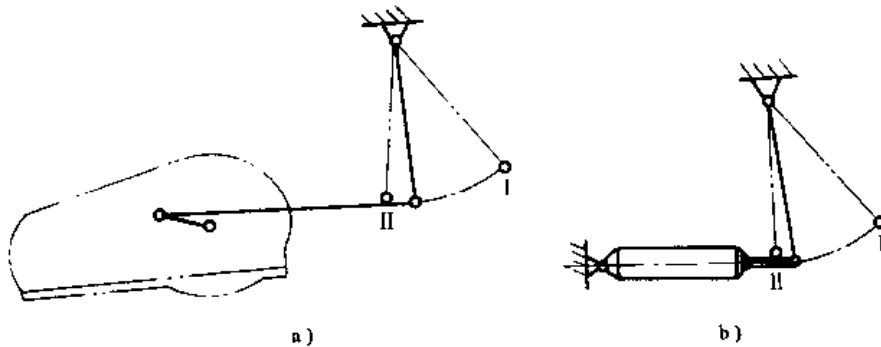


图 47.3.38 摆动机构方案比较

### 5.5 机构类型创新和变异设计

在机构构思和设计时要凭空构想出一个能达到预期动作要求的新机构,往往非常困难。采用机构类型创新和变异的创新设计方法,则是借鉴现有机构的运动链类型,进行类型创新或变异创新来得到新的机构类型,满足新的设计要求。这种方法的基本思想是:将原始机构用机构运动简图表示,通过释放原动件、机架,将机构运动简图转化为一般化运动链,然后按该机构的功能所赋予的约束条件,演化出众多的再生运动链与相应的新机构。

下面以越野摩托车尾部悬挂装置的创新设计为例说明机构类型创新的步骤和方法。机构类型创新设计的程序如图 47.3-39 所示。

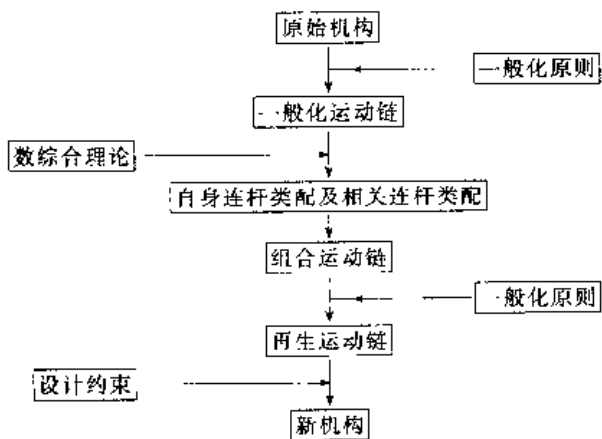


图 47.3-39 机构创新设计流程

#### (1) 原始机构

图 47.3-40 所示为越野摩托车尾部悬挂装置的原始机构。图中,1 为机架,2 为支承臂,3 为摆动杆,4 为浮动杆,5、6 分别为减震器的活塞和气缸。

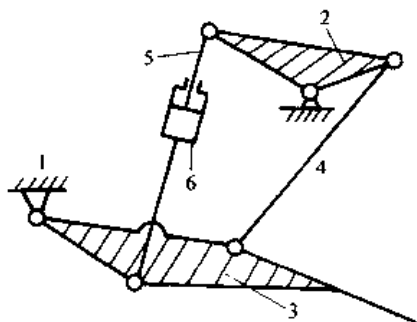


图 47.3-40 越野摩托车后部原始机构

#### (2) 一般化运动链

一般化运动链是只有连杆和转动副的运动链。图 47.3-41 为图 47.3-40 所示的原始机构的一个一般化运动链。将原始机构简图抽象为一般化运动链的一般化原则为

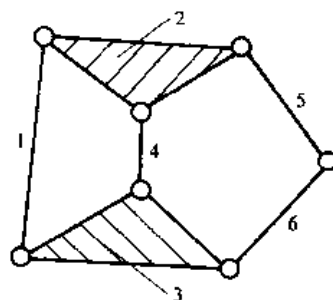


图 47.3-41 一般化运动链

- 1) 将非连杆形状的构件转化为连杆;
- 2) 将非转动副转化为转动副;
- 3) 机构的自由度应保持不变;
- 4) 各构件与运动副的邻接应保持不变;
- 5) 固定杆的约束予以解除,使机构成为一般化运动链。

图 47.3-42 所示为两种具有六杆、7 个运动副的一般化运动链。

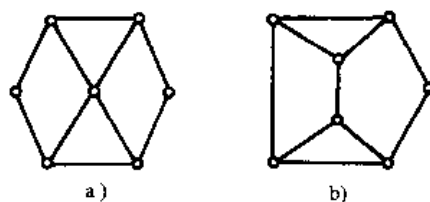


图 47.3-42 六杆、7 个运动副的一般化运动链

#### (3) 设计约束

对越野摩托车尾悬挂装置先订出几个设计约束,作为新机构类型创新的依据。这些设计约束为

- 1) 必须有一固定杆作为机架;
- 2) 必须有吸震器;
- 3) 必须有一个摆杆安装摩托车后轮;
- 4) 固定杆、摆杆和吸震器必须是不同的构件。

(4) 具有固定杆的特殊运动链  
若以 G 表示固定杆、S<sub>0</sub> 表示摆杆、S<sub>1</sub>S<sub>2</sub> 表示吸震器。

根据设计约束,对图 47.3-42 所示的两套运动链,参照有关理论算法,可得出图 47.3-43 的具有固定杆的特殊运动链的 10 种类型。

#### (5) 新机构

对于实际设计问题,其约束是多变的。对于本例的悬挂装置,如果没有实际约束,则图 47.3-43 的所有类型是可行的。

若我们再定义该机构创新设计的约束条件为:摆杆与固定杆相连。按此约束条件,图 47.3-43 中能够满足设计要求的可行设计方案只有 6 个,即图 47.3-43a、b、d、f、h、i 这 6 种合适的新机构,如图 47.3-44 所

示。

很明显,机构类型创新只解决机构的型、数综合问

题,要实现某一动作还应进行机构的尺度综合。只有这样,才能算是完成机构的创新设计。

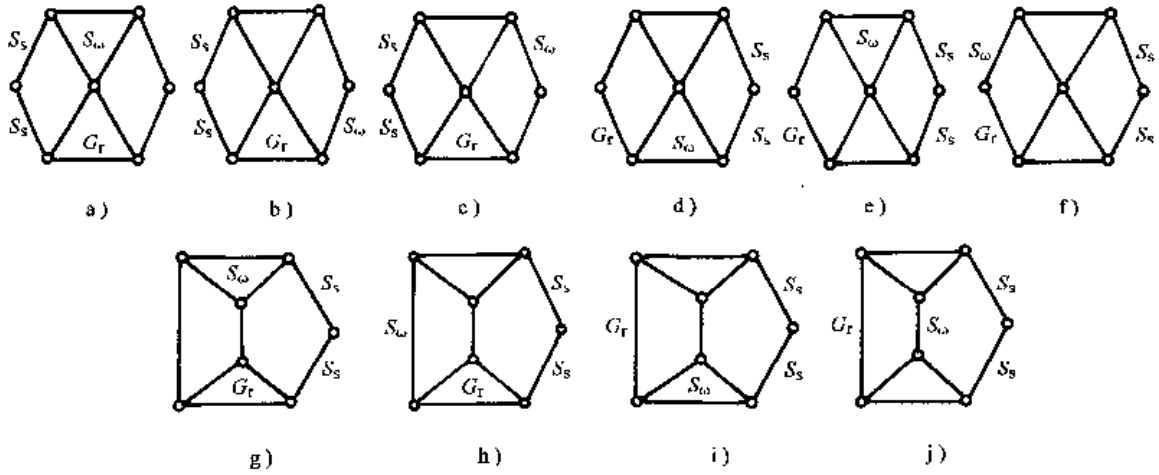


图 47.3.43 特殊运动链类型

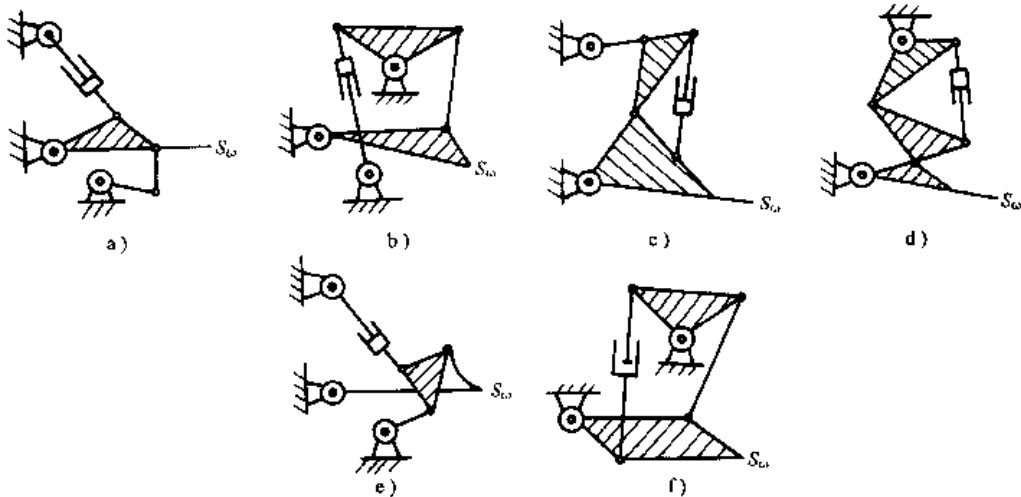


图 47.3-44 运动链合适的新机构

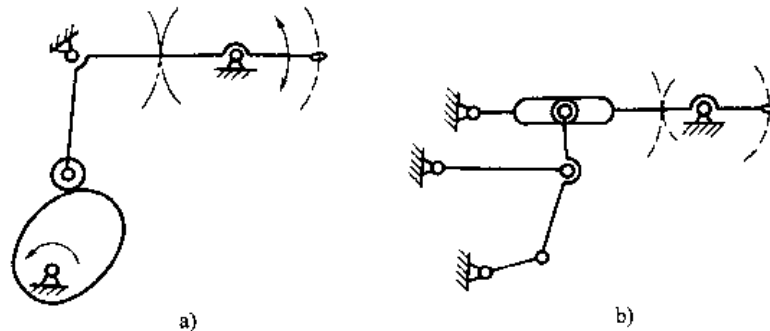


图 47.3-45 挑线机构的类型变换

a) 原挑线机构 b) 变换后的挑线机构

### 5.6 机构类型变换法

机构类型变换法是采用结构不同的机构替代原有机构而使两者的输入、输出的运动特性相同或等效的

创新设计方法。

类型变换法可以较好地解决专利替代、改善性能、简化加工工艺等问题。

图 47.3-45a 为某缝纫机的挑线机构,它的挑线轨



迹主要由摆动从动件凸轮机构来决定，通过设计合适的凸轮轮廓曲线实现供线和收线的要求。图 47.3-45b 是类型变换后的挑线机构，它的挑线轨迹主要由四连杆机构的某点的连杆曲线来决定，通过选择合适的铰链四杆机构的参数来实现供线和收线的要求。实验证明图 47.3-45b 的机构既避开了专利，又使噪声下降、制造工艺简化。

以上 7 种机构创新设计方法，有利于设计人员对执行机构的创新，寻求新颖的执行机构，提高机械运动系统设计的创造性程度。

### 6 机构选型

机械运动系统方案的构思、设计、拟定是机械创新设计的主要内容。执行机构选型就是按功能和执行动作的要求选择可行的各种执行机构类型。因此，机构选型是机械运动系统设计的重要步骤。机构型式的选择

是否合适，直接关系到机械运动系统的先进性、适用性和可靠性。

从广义上说，机构选型应包括选择已有的合适机构型式，或者创造前所未有的机构型式。判别机构型式的好坏主要看是否实现执行动作的形式和规律。机构选型时要充分依靠设计者的经验和直觉知识，但是还应借助机构选型的基本方法和主要规律。

虽然这些方法、规律永远不可能代替设计者的创造力，但可以扩大设计者的知识面，运用创造技法提高设计效率，在较高的设计水平上提供选择机构的办法。

#### 6.1 按运动形式选择机构

常见的机器工艺动作所要求的运动形式有：单向转动、摆动、移动；双向转动、摆动、移动；摆动至极限位作停歇运动。为了形象直观地加以描述，表 47.3-1 列出了常用的运动形式与表达符号。

表 47.3-1 常用运动形式与表达符号

运动形式	连续运动			间歇运动			极限位停歇	
	转动	摆动	移动	转动	摆动	移动	摆动	移动
单向								
双向								

设计者要解决的问题，是把原动机的运动形式和运动参数转换为各种各样的执行构件的运动形式和运动参数。一个原动机往往要驱动多个执行构件动作，因此在原动机与执行构件之间必须采用具有不同功能的机构来进行运动参数和运动形式的转换，以实现执行构件的预期动作。由此看来，对机器运动变换必须进行功能分析，即对原动机和执行构件所形成的传动链两端的运动形式和运动参数进行分析研究。只有考虑到在运动传递过程中轴线位置的配置等功能要求，才能确定各中间机构的基本功能。如由转动转换成移动、摆

动、间歇转动等。将基本功能用符号来表示，并配上符合该功能的机构型式，这样设计者可以较方便地进行选择或在此基础上进行机构的变异和创造。

表 47.3-2 列出了常见的运动转换基本功能及实现这些基本功能的机构。表中基本功能用两个矩形框中的符号表示，表示由左框中的运动形式转换为右框中的运动形式。

由 47.3-2 可知，实现某一运动转换基本功能的机构型式有多种。因此把这些机构按运动的传递顺序组合起来构成的运动方案也有很多种。

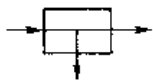
表 47.3-2 常见运动转换基本功能、代表符号和匹配机构

名称	符号	功能载体
转动变为双向摆动		曲柄摇杆机构，摆动从动件凸轮机构 曲柄摇块机构，电风扇摇头机构 摆动导杆机构，曲柄六连杆机构
转动变为单向间歇转动		槽轮机构，平面凸轮回歇机构 不完全齿轮机构，圆柱凸轮分度机构 针轮回歇转动机构，蜗杆凸轮分度机构 偏心轮分度定位机构，内啮合星轮回歇机构

(续)

名 称	符 号	功 能 载 体
转动变为单侧间歇摆动		摆动从动件凸轮机构, 连杆曲线间歇摆动机构 曲线槽导杆机构
转动变为双向间歇摆动		六连杆机构两极限位置停歇摆动机构 四连杆扇形齿轮双侧停歇摆动机构
转动变为实现预定轨迹		行星轮直线机构, 联动凸轮机构 铰链五杆机构, 铰链六杆椭圆轨迹机构 连杆凸轮组合机构, 行星轮摆线正多边形轨迹机构
摆动变为单向间歇转动		棘轮机构, 摩擦钢球超越单向机构
转动变为单向直线移动		齿轮齿条机构, 螺旋机构, 带传动机构, 链传动机构
转动变为双向直线移动		曲柄滑块机构, 六连杆滑块机构 移动从动件凸轮机构, 不完全齿轮齿条机构 连杆凸轮组合机构, 正弦机构
转动变为单侧间歇移动		连杆单侧停歇曲线槽导杆机构 移动凸轮间歇移动机构 行星内摆线间歇移动机构
转动变为双侧间歇移动		不完全齿轮齿条往复移动间隙机构 (用于印刷机) 不完全齿轮移动导杆间歇机构, 移动从动件凸轮机构 八连杆滑块上下端停歇机构 (用于喷气织机开口机构)
运动缩小		齿轮传动机构, 谐波传动机构, 链传动机构 行星传动机构, 摆线针轮传动机构, 摩擦轮传动机构
运动放大		蜗轮蜗杆机构, 带传动机构, 螺旋传动机构 流体传动机构, 连杆机构
运动合成		差动螺旋机构 (用于测微机、分度机构、调节机构、夹具等) 差动轮系, 差动连杆机构, 二自由度机构
运动分解		差动轮系, 二自由度机构
运动换向		惰轮换向机构, 棘轮机构 滑移齿轮换向机构, 摩擦差动换向机构 行星齿轮换向机构, 离合器锥齿轮换向机构
运动轴线变向		锥齿轮传动, 交叉带传动 螺旋齿轮传动, 蜗杆传动, 单方向联轴器, 圆柱摩擦轮传动
运动轴线平移		圆柱齿轮传动, 带传动, 链传动 平行四边形机构, 双向联轴器, 圆柱摩擦轮机构

(续)

名称	符号	功能载体
运动分支		齿轮系·带轮系, 链轮系

注: 参考资料: 徐灏主编,《机械设计手册》第2卷: 8. 机构及机械系统设计。辛一行等主编,《现代机械设备设计手册》第1卷, 第8篇: 机构及其系统设计。

## 6.2 按执行机构的功用选择机构

有时, 机器执行动作的功用很明确。如夹锁紧、分度、定位、制动、导向等功用。设计者可以遵照这些功用, 先查阅有关机构手册, 熟悉相应的夹锁紧机构、分度机构、定位机构、制动机构和导向机构等内容。然后进行分析对比, 选择确定一种适用于所设计机器的机构。即使选择不到合适的机构型式, 也能达到开阔思路, 借鉴启发的作用, 可在已有机构型式的基础上增加辅助机构或组合成新机构等方法, 实现执行动作的功能要求。

## 6.3 按不同的动力源形式选择机构

常用的动力源有电动机、气液动力源、直流电动机等。机构选型时应充分考虑动力源情况和生产情况, 当有气液源时常用气动、液压机构, 这样可以简化机械结构, 省去传动机构与运动转换的机构。特别对具有多个执行构件的工程机械、自动生产线和自动机等, 更应优先考虑。

## 6.4 按先易后难选择机构

表 47-3-2 列出了多种运动转换基本功能。每种基本功能可由多种机构型式来实现, 其中大部分属于较简单的基本机构。实际上, 每种运动转换基本功能还可匹配多种机构组合和组合机构。设计者如何从众多的机构型式中合理选择呢? 通常的选择顺序是: 先选基本机构, 再选机构组合, 最后选组合机构。

例如, 实现转动转换为摆动的功能。基本机构: 曲柄摇杆机构; 机构组合: 曲柄摇杆机构串连-导杆机构; 或串连-铰链四杆机构等, 都能满足这样的转换功能。设计者可先选基本机构, 只有当曲柄摇杆机构不能满足摆幅要求, 或不能满足动力特性要求时, 再考虑选择曲柄摇杆机构串连导杆机构的组合型式。以获得从动件更大的摆角幅度。所以机构选型, 必须对各类机构的运动特性有深入的理解, 才能选得准、选得好。

## 6.5 选择机构及其组合安排时应考虑的主要要求和条件

### (1) 运动规律

执行构件的运动规律及其调节范围是机构选型及机构组合安排的基本依据。

### (2) 运动精度

运动精度的高低对机构选型影响很大。例如, 对运动速度和运动时间要求很高时, 就不宜采用液压和气压传动; 如果对运动精度要求不高, 可采用近似直线运动代替直线运动, 用近似停歇来代替停歇。这样, 可使所选机构结构简单, 易于设计、制造。

### (3) 承载能力与工作速度

各种机构的承载能力和所能达到的最大工作速度是不同的, 因而需根据速度的高低、载荷的大小及其不同的特性来选用合适的机构。

### (4) 总体布局

原动机与执行构件的工作位置、传动机构与执行机构的布局要求等是机构选型和组合安排必须考虑的因素。要求总体布局合理、紧凑, 使机械的输出端尽可能靠近输入端, 这样可省去不必要的传动机构。

### (5) 使用要求与工作条件

使用单位所提出的生产工艺要求, 生产车间的条件、使用和维修要求等, 均对机构选型和组合安排有很大影响。

## 7 动作解法库的建立

为了易于寻求最佳动作载体, 已有不少手册按机构实现动作的形式分类列出相应的执行机构。但是这种按动作的形式分类列出的执行机构还是比较粗糙的。我们必须建立起具有更多信息要求的动作解法库。可以把机构的输入-输出类型、机构输出运动的基本特性(如运动类型、运动轴线、速度方向、运动的连续性、速率变化特性等)以及机构的变化特性(如工作性能、动力性能、经济性、结构尺寸等)充分反映出来, 列出分类动作解法库, 便于设计者选择最佳动作载体。

## 第4章 机械运动系统的协调设计

### 1 机械运动系统的基本构成

机械产品从整体需要出发,通常可以有下列一些子系统:动力系统、传动系统、执行系统和操作及控制系统等。我们常常把传动系统和执行系统统称为机械运动系统。如图 47.4-1 表示了机械产品的基本组成图。对于机械产品来说机械运动系统的构成是比较复杂的,它的设计富有创造性。因此,往往作为机械产品设计的重点。

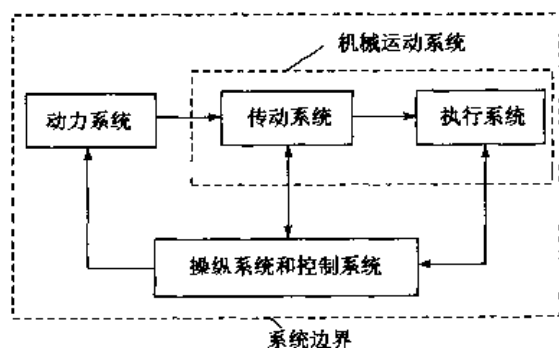


图 47.4-1 机械产品的组成

#### 1.1 传动系统

在一般的机械产品中是把动力机的动力和运动传递给执行系统的中间机械装置。

传动系统主要起变速、换向、传递运动等作用。具体来说,传动系统有如下一些功能:

1) 减速或增速:将动力机的速度降低或增高,以适应执行系统的工作需要;

2) 变速:进行有级或无级变速,以满足执行系统工作速度变化的要求;

3) 改变运动规律或运动形式:把动力机的等速转动转变为按某种规律变化的转动、摆动或移动,其中包括间歇转动和间歇移动等。也可以通过传动系统将动力机输出的转动方向改变成相反转动方向。

4) 传递动力:把动力机输出的动力传递给执行系统,供执行系统完成预定工艺动作过程所需的驱动力矩或驱动力。

机械产品的传动系统按传动比或输出速度是否变化可分为固定传动比和可调传动比系统;按动力机驱动执行机构或执行构件的数目可分为独立驱动、集中驱动和联合驱动的传动系统。

#### 1.2 执行系统

执行系统是机械产品中用以直接完成预期工艺动作过程的子系统。它利用机械能改变工作对象的形态或搬移工作对象。

##### (1) 执行系统的基本组成

执行系统基本上是以一系列执行机构所组成。执行构件的动作是由执行机构来产生的。一系列执行构件所产生的工艺动作过程完成了机械产品的总功能。例如,完成夹持、搬运、转位、间歇动作等等的运动动作。又如完成喷涂、冲压、切削等的改变工作对象的形态的工作。

执行构件一般是执行机构中的一个构件,它的运动形式和运动精度取决于整个机械系统的工作要求。在选择执行机构时必须考虑这些工作要求。

##### (2) 执行系统的功能

执行机构的功用是把传动系统传递过来的运动与动力进行必要的转换,以满足执行构件完成功能的需要。

执行机构变换运动的类型主要是将转动变换为移动或摆动,或则相反将移动或摆动变换为转动。而从执行构件的具体运动方式来看,可分为将连续运动变换为不同形式的连续运动或间歇运动。

执行系统是在几个执行机构所带动的执行构件协调工作下完成机械系统的任务。虽然工作任务多种多样,但执行系统的功能主要有以下几种:夹持、搬运、输送、分度与转位、检测、施力等等。机械系统的任务只是由执行系统有限的功能来完成。

执行系统工作性能的好坏,直接影响整个系统的性能,从机械系统概念设计要求来看,主要是考虑运动精度和动力学特性等要求。

### 2 机械运动系统设计

#### 2.1 机械运动系统的基本内容

机械运动系统又可称为机构系统,是由一系列传动机构和执行机构组成。

机械运动系统是机械系统的重要内容。机械系统设计的好坏,将直接影响到整个机械系统技术性能的优劣、机械系统结构的繁简、机械系统制造成本高低、以及操作的难易。机械系统系统设计的好坏,关键在于

机械运动系统方案设计的优劣。如果在机械运动系统方案设计时考虑欠缺,就会造成先天不足,即使在后续设计阶段采取一些补救措施,也是治标难治本,整个运动系统的总功能难以得到根本性改善。所以说,在进行机械运动原理方案设计时,必须进行以下几个方面的研究和探讨。

1) 根据机械的功能要求,注意市场的调查和预测,通过需求分析,确定出一种能满足新的需求、具有竞争力的新型机械型式。

2) 要尽量采用先进技术,推陈出新,既不脱离实际、盲目追求先进,也不能因循守旧墨守成规,要在继承的基础上,大胆创新。

3) 对尚无确切把握的新方案,应进行必要的试验、研究和验证。对于试验中所出现的问题和不足,必须采取一些完善和修改的措施后,方可着手进行正式设计。

4) 在处理机械的使用要求与制造要求的关系时,

应该在满足机械的使用要求(亦即保证实现机械的功能)下考虑机械的制造问题。优良的制造和装配质量是提高机械使用性能、工作可靠性和使用寿命所必不可少的条件。

机械运动方案的设计应考虑以下一些基本内容。

#### (1) 功能与应用范围的确定

一台机器的功能是要完成某一工艺动作过程,一个工艺动作过程又可以分解为若干个工序。

例如,制盒工艺为纸盒成型—物料充填—封口,它的工艺动作可以分解为图 47.4-2 所示的 9 个工序:纸盒料坯从贮存器中送出①,纸盒成型②,纸盒一端盒口闭合③,纸盒一端盒口封口④,纸盒翻转 90°⑤,盒子送进⑥,物料充填纸盒⑦,另一端盒口闭合⑧,另一端盒口最后封口⑨。工艺动作过程的分解与工作原理、功能和应用范围密切相关,也与实现各工序动作的方法有关。采用简单实用的机构来实现工序动作,是每一个设计师应努力达到的目标。

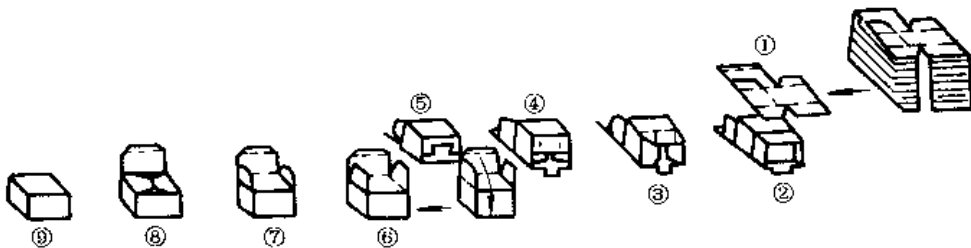


图 47.4-2 纸盒成型的工艺动作过程分解的工序

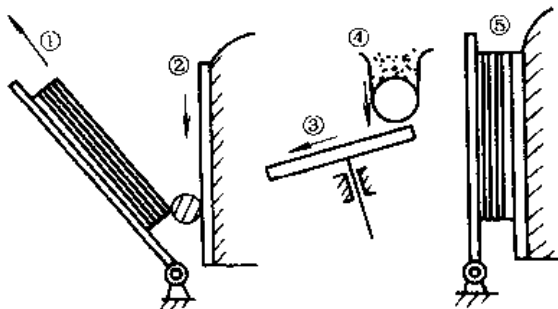


图 47.4-3 平版印刷机工艺动作过程

图 47.4-3 所示为平版印刷机所需完成的 5 个工序:取出已印刷好的纸张①,墨辊向印版上滚刷油墨②,墨盘间歇转动一个位置,使油墨匀布于墨盘,以便墨辊滚过墨盘时得以均匀上墨③,将油墨容器内的油墨源源不断供应给油盘④,空白纸张合在印版上完成印刷⑤。如果是取出机上已加工印件,那么平板印刷机只需完成后面的②、③、④、⑤工序,可以用 4 个机构分别完成。

工业平缝机的功用与普通家用缝纫机没有多大的差别,对于工业平缝机,它所要完成的工艺动作过程一

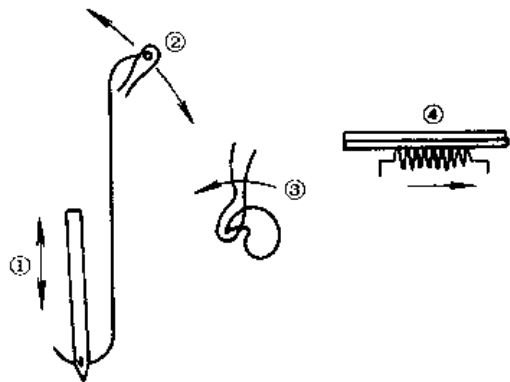


图 47.4-4 工业缝纫机工艺动作过程

般可以分解为 4 个工序,如图 47.4-4 所示。

① 刺料: 缝纫机针刺进和退出缝料。机针刺入缝料后形成线圈,为钩线创造条件;机针退出缝料以便收线和送料。

② 挑线: 完成缝纫过程中的供线和收线的要求,使缝料上线迹良好,不产生断线和松线。

③ 钩线: 完成钩住面线线圈,使面线与底线交织起来。

④ 送料: 在缝料完成一个线迹后,使缝料向前进

送一个针距,以便进行下一个线迹的缝纫。送料机构的运动轨迹,要求在送料部分平直。

以上所列举的 3 种机器,由于所完成的功能比较单一,因此工艺动作过程的分解一般比较简单。实际上,在创新设计一台机器之前,对工艺动作分解进行种种构思,就可得到不同的机械运动方案设计。而且,用上述方法所得到的机器是多功能类型的。

在设计一台完成特定工艺动作过程的新机器时,还可以按分解后的工序,用若干台机器分别完成一系列工序的单功能机器组合。这种做法在有些情况下是合理的,但是对不少场合采用单功能机器组合是不现实的。在不少实用机器设计时,往往设计成多功能类型,也就是用一台机器完成一系列工艺动作(即多道工序)。

多功能类型的机器有如下的优点:

1) 可以省去单机之间的连接和输送装置,减少动力机和传动变速器,有助于简化机械结构,降低成本,提高劳动生产率;

2) 可以缩小机器占地面积;

3) 减少操作人员,改善劳动条件;

4) 有利于机器工作性能的提高。

确定机器的功能与应用范围,必须考虑机器的可靠性和适应性。

#### 可靠性

在一般情况下,机器功能增多,必然使工艺动作过程的工序增多,故障发生的可能性一般也相应增大。为了提高一台机器的可靠性,一般可以将工艺动作过程所分解出的几个工序分成若干组,每组工序用一台单机来完成,整个工艺动作过程用几台单机来完成。但是这种做法,使各单机之间连接比较复杂,增加机器组合的故障。有不少机器,例如前面所述的平版印刷机、工业平缝机,它们工艺动作全过程的各个工序在一台机器上实现更为合理、可靠,因此就没有必要用几台单机来完成。从提高机器的可靠性考虑,实现同一机器功能时,尽可能采用动作简单、工序数少的工艺动作过程。

#### 适应性

任何机器它的应用范围都是有限的,不可能是万能的。例如,盒式包装机对于盒子的大小、充填物的体积和状态都有一定的规定,并不是一概适应的。又如工业平缝机对于缝料厚度、缝料质地等的适应性,也有一定范围,使用者不可以随心所欲。机器的功能愈多,机器的结构也就愈复杂。因此,对于将机器功能分解后用几台单机来完成的机器组合,为了增强其适应性,比较合理的办法是根据用户要求,灵活地增减或改装某些组合部件,以扩大应用范围,满足用户的不同需要。

机器的功能和应用范围,还与产品(加工对象)批量及品种规格有关。对于产品批量大、品种规格稳定的产品,在设计机器时应致力于提高机器的生产率、机器自动化程度,一般宜采用专机类型。对于批量中等、品种规格需要调换的产品,一般可通过调整或更换有关部件来适应各批生产的需要,此时宜采用多用机类型。对于批量小、品种规格经常变化的产品,应尽量扩大机器的应用范围,以利增加机器制造的批量和减少设备投资,此时宜采用通用机类型。

#### (2) 工作原理和工艺动作分析

根据机械功能要求,首先应选择机械的工作原理。机械完成同一种功能,可以应用不同的工作原理来实现。例如,印刷机可以采用平版印刷工作原理,也可采用轮转式印刷工作原理。平版印刷机工作原理是在平放的铜锌字版上涂油墨,待印件覆在铜锌字版上完成印刷工艺,其工作原理与盖图章动作相类似。轮转式印刷机工作原理是将铜锌字版装在圆筒表面,随圆筒连续转动,待印纸张相应连续移动,与圆筒贴合,完成印刷工艺。轮转式印刷的工作原理与盖图章动作大相径庭,它的生产率大大提高。平版印刷机和轮转式印刷机相比较,它们之间的结构、造价、生产率有很大差别。再如,螺栓的螺纹加工,有车削、套螺纹、滚压和拉丝。车削是采用螺纹车刀在车床上切削加工;套螺纹是采用螺纹板牙在机床上(或手工)加工;滚压是采用一副滚螺纹轮在滚螺纹机上加工;搓螺纹是采用一副搓螺纹板在搓螺纹机上加工。后三者特殊的工艺动作过程所采用的机器与车床也有明显不同。

工作原理的选择与产品加工的批量、生产率、工艺要求、产品质量等等有密切关系。在选定机器的工作原理时,不应墨守成规,而是要进行创新构思。构思一个优良的工作原理可使机器的结构既简单又可靠,动作既巧妙又高效。

工作原理与工艺动作过程的设计与构思,还应充分考虑被加工对象的材料特性、所需达到的工艺要求和生产率等。

工作原理与工艺动作过程的设计与构思,还要考虑机器的工位数和工位转移过程中采用间歇运动还是连续运动。工位多时执行机构要分散布置。工位转移过程中采用间歇步进运动,可使执行机构的运动形式构造比较简单,而连续运动会使执行机构复杂化。

工艺动作过程的确定,还应考虑工艺程序和工艺路线。工艺程序是指完成各个工艺动作的先后顺序。例如图 47.4-5 所示折叠裹包程序为

物品到位→包装纸到位→折叠包装纸的上、下、前三面→后面包装纸折叠→两侧包装纸折叠→两侧下面裹包→两侧上面裹包→包装后的物品输出。

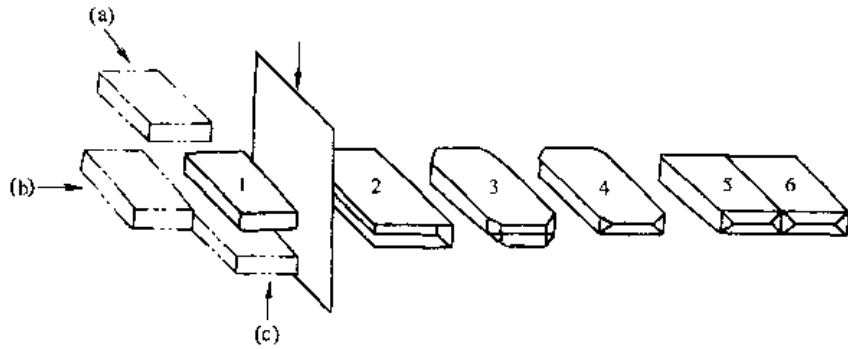


图 47.4-5 折叠式包装工艺动作过程分解

工艺程序的编排对于机器结构简化、工作可靠性能起较大的作用。在上述例子中对包装纸的上、下、前三面的折叠只采用了将物品向右推进的机构，利用物品上下的挡块形成上、下、前三面包装纸的折叠(如图 47.4-6 所示)，这种方式结构简单、工作可靠。

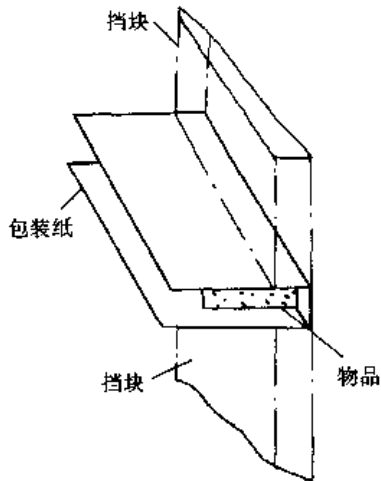


图 47.4-6 折叠式包装机构

工艺路线是指参与加工的物料的供送路线、加工物料的传送路线以及成品的输出路线。完成同一工艺程序的工艺路线可以多种多样，常见的有：

- 1) 直线型 物料的运动路线为一直线，根据运动方向又可分为立式和卧式两种，如图 47.4-5 所示。
- 2) 阶梯型 物料的运动路线兼有垂直和水平两方向，如图 47.4-7 所示。
- 3) 圆弧型 物料的运动沿圆弧轨迹，如图 47.4-8 所示。
- 4) 组合型 物料的运动既作圆弧运动，又作直线运动，如图 47.4-9 所示。

选择工艺路线时，必须分析比较各种工艺路线的特点，既要考虑它对机器生产率、执行机构数目和运动要求的影响，又要考虑它对机器外形、操作条件以及组成加工生产线等方面的影响。只有对各种不同工艺路

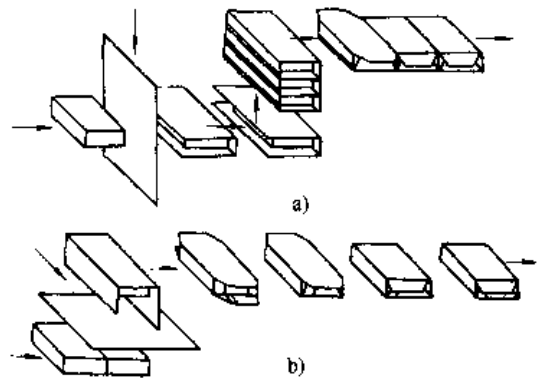


图 47.4-7 阶梯型工艺路线

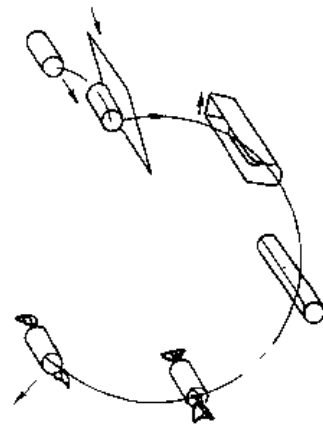


图 47.4-8 旋转型工艺路线

线的方案加以认真详细的分析，才有可能找到最优方案。很显然，机器的工艺路线对设计机械运动方案影响很大。

### (3) 执行构件的运动要求和执行机构的布局

根据机器的工作原理和工艺动作要求，可以进行分析、构思确定执行构件的数目和运动要求。执行构件分为固定的和运动的。例如，在不少包装机械中，成型器和固定折纸板等均属固定式，它们的形状依据工艺要求来确定。大多数执行构件是按一定的工艺运动要求来确定它的运动规律，包括运动的形式、行程大小、

动停时间和运动速度等。执行构件的运动规律一旦确定后,就可着手于选定执行机构。

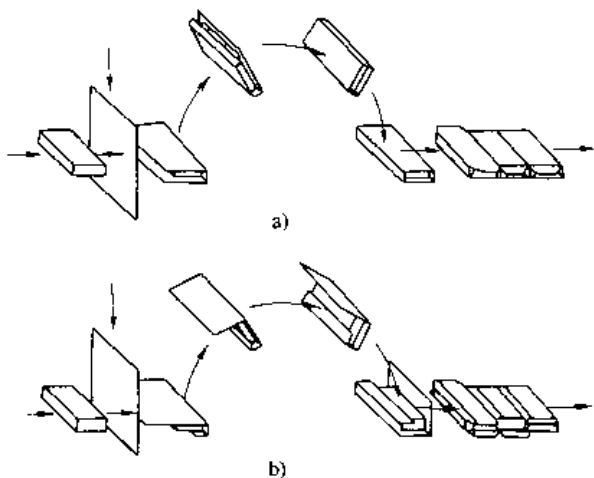


图 47.4-9 组合型工艺路线

执行机构布置的步骤主要分为两步,即根据工艺路线图将各个执行构件布置在预定的工作位置上;布置执行机构的原动件,对于气液压传动主要安排气缸和油缸的位置,对于机械传动则是安排机构输入构件凸轮、齿轮、曲柄等原动件的位置。

执行机构的原动件布置要遵守两条原则:一是使原动件应尽可能接近执行构件,这样可使执行机构简单紧凑并尽可能减小其几何尺寸;二是使原动件尽可能集中布置在一根轴上或少数几根轴上,这样可以使整个传动系统大为简化,同时便于对机器进行调试和维修。为了使执行机构能同步运动,各原动件所在的几根原动轴转速应保持等速或定速比。

#### (4) 机器工艺过程的设计原则

机器工艺过程设计的总要求是保证产品加工的质量,具有较高的生产率,机器的结构力求简单,操作和维修方便,制造成本低和维护费用小等等。为了达到上述总要求,机器的工艺过程一般应遵循以下几个设计原则:

1) 工序集中原则 工序集中原则是指工件在一个工位上一次定位装夹,采用多刀、多面、多个执行构件动作同时完成加工工序,以完成工件的工艺要求。工序集中原则使加工质量容易保证,机器的生产率也较高。

2) 工序分散原则 工序分散原则是将工件的加工工艺过程分解为若干工序,并分别在各个工位上用不同的执行机构进行加工,以完成工件的工艺要求。由于工序分散,执行机构完成每一工序的动作较为简单,这样可以使机器生产率有较大的提高。

工序集中原则和工序分散原则从表面上看是矛盾的,其实又是统一的,因为依据实际情况,工序能集中就尽量集中,工序集中有困难就采取分散。集中是为了

提高机器生产率,分散也是为了提高机器生产率。两个原则为了同一目的,只是在不同场合采用不同方法而已。

3) 各工序的工艺时间相等原则 对于多工位自动机工作循环的时间节拍有严格的要求,一般将各工位中加工时间最长的一道工序的工作循环作为自动机的时间节拍。为了提高机器的生产率,可以将工艺时间较长的工序,采取适当的措施,如提高这一工序的工艺速度或将这一工序实行再分解等等。

4) 多件同时加工原则 多件同时加工原则是指在同一自动机上同时加工几个工件,也就同时采用相同的几套执行机构(或执行构件)来加工多个工件。这样可以使机器的生产率成倍提高。

5) 减少机器工作行程和空程时间 设计工艺过程时,为了提高机器的生产率,在不妨碍各执行构件正常动作和相互协调配合的前提下,尽量使各执行机构的工作行程时间互相重叠、工作行程时间与空行程时间互相重叠、空行程时间与空行程时间互相重叠,从而缩短工件加工循环的时间。

#### (5) 机器生产率的分析与计算

在进行机械运动原理方案设计时,设计人员应在保证工件加工质量的前提下,力求使机器具有较高的生产能力。

1) 机器的生产率 机器的生产率是指机器在单位时间内生产的产品数量。一个工件所需的加工时间  $T$  为:

$$T = t_{\text{工作}} + t_{\text{空程}}$$

式中  $t_{\text{工作}}$  —— 工作行程所需的时间;

$t_{\text{空程}}$  —— 空行程所需的时间。

因此,机器的生产率  $Q$  为:

$$Q = \frac{1}{T} = \frac{1}{t_{\text{工作}} + t_{\text{空程}}} \quad (47.4-1)$$

式(47.4-1)所表示的机器生产率是机器调整到正常工作状态下的单位时间内加工的产品数量,称之为机器的理论生产率。

如果机器没有空行程,即空程=0,那么它的生产率称之为机器的连续生产率,以  $K$  表示:

$$K = \frac{1}{t_{\text{工作}}} \quad (47.4-2)$$

式 47.4-2 表示机器连续工作,没有空行程或空行程时间,同时各执行机构的工作行程时间完全重合。

机构的理论生产率  $Q$  又可表示为:

$$Q = \frac{K}{K \cdot t_{\text{空程}} + 1} \quad (47.4-3)$$

$$\eta = \frac{1}{K \cdot t_{\text{空程}} + 1} \quad (47.4-4)$$

$\eta$  称为生产率系数。由此可见生产率系数表示机器的



理论生产率与连续生产率之比,即

$$\eta = \frac{Q}{K} \text{ 或 } \eta = \frac{t_{\text{工作}}}{T}$$

因此,生产率系数  $\eta$  表示机器在时间上的利用程度,亦即反映工艺过程连续化程度。

2) 机器生产率的分析 为了便于分析,将式(47.4-3)和式(47.4-4)用图形来表示。图 47.4-10 是具有不同的  $t_{\text{空程}}$  值的生产率系数曲线。图 47.4-11 是具有不同的  $t_{\text{空程}}$  值的机器生产率曲线。图中曲线 I、II、III 分别为不同的  $t_{\text{空程}}$  值时机器的  $\eta$  与  $K$  及  $Q$  与  $K$  的相互关系。其中  $t_{\text{空程}}^{\text{I}} < t_{\text{空程}}^{\text{II}} < t_{\text{空程}}^{\text{III}}$ 。由此得出如下的分析结果:

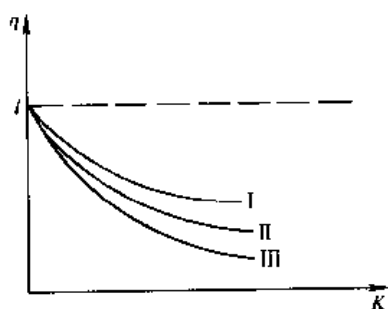


图 47.4-10 不同  $t_{\text{空程}}$  值的生产率系数

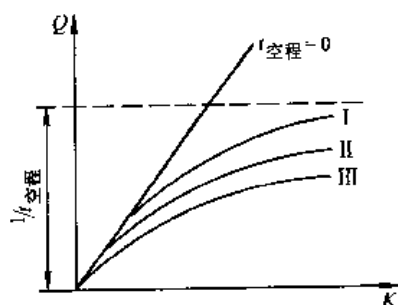


图 47.4-11 不同  $t_{\text{空程}}$  值的生产率

(a) 机器的生产率系数值的变化范围为  $0 < \eta < 1$ 。理论生产率随连续生产率的提高而增大;生产率系数将随连续生产率的提高而减少。

(b) 当  $K$  为定值时,理论生产率  $Q$  将随空行程时间  $t_{\text{空程}}$  的减少而增加,其极限值为:

$$Q = \lim_{t_{\text{空程}} \rightarrow 0} \frac{K}{K \cdot t_{\text{空程}} + 1} = K$$

(c) 在  $t_{\text{空程}}$  为定值时,理论生产率将随连续生产率的  $K$  的提高而增加,当  $K \rightarrow \infty$  时,  $Q$  的极限值为:

$$Q = \lim_{K \rightarrow \infty} \frac{K}{K \cdot t_{\text{空程}} + 1} = \frac{1}{t_{\text{空程}}}$$

(d) 在连续生产率  $K$  较低时,  $t_{\text{空程}}$  对  $Q$  的影响较小,在连续生产率  $K$  较高时,  $t_{\text{空程}}$  对  $Q$  的影响显著。

由以上分析可以得出结论:在对  $K$  值较小的机器进行设计时重点设法提高连续生产率  $K$ ;在对  $K$  值较

大的机器进行设计时,应重点设法减小  $t_{\text{空程}}$ ,才能有效地提高机器的生产率。

(e) 没有空行程时间(即  $t_{\text{空程}} = 0$ )的机器,就是连续作用式的机器,这时理论生产率等于连续生产率。一般自动机械很少有这种情况。

3) 机器的实际生产率 机器的实际生产率往往低于机器的理论生产率,这是因为存在由于种种原因所造成的停车时间。所以机器的实际生产率  $Q_{\text{实}}$  可以用下式来表示:

$$Q_{\text{实}} = \frac{1}{t_{\text{工作}} + t_{\text{空程}} + t_{\text{损}}}$$

式中  $t_{\text{损}}$ ——机器工作循环以外的损失时间。

图 47.4-12 是机器的理论生产率  $Q$  和实际生产率  $Q_{\text{实}}$  同连续生产率  $K$  之间的关系曲线。图中曲线 I 表示为连续式机器(即  $t_{\text{空程}} = 0$ )的生产率变化曲线;曲线 II 表示机器的理论生产率曲线;曲线 III 代表机器的实际生产率变化曲线。

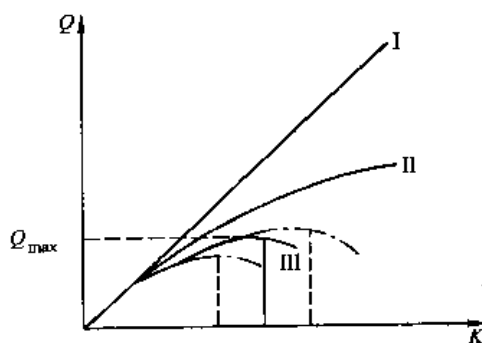


图 47.4-12 机器理论生产率与实际生产率关系

## 2.2 机械运动系统的集成设计

机械运动系统的设计目标是实现所要求的功能,满足市场需求。机械运动系统的设计应包括:确定功能要求、选择工作原理、构思工艺动作过程、分解工艺动作过程为若干执行动作、创新或选定执行机构、执行机构系统的集成等。

机械运动系统的集成设计是指如何将确定的执行机构按工艺动作过程要求进行系统集成,使机械运动系统中各执行机构之间达到运动协调、使系统整体功能达到综合最优。

下面介绍一下机械运动系统集成设计的基本原理:

### (1) 合理地分解工艺动作过程

在机械总功能和工作原理确定之后,工艺动作过程的类型一般是比较有限的。例如,图 47.4-5 所示折叠式包装工艺动作过程,只是存在被包物件开始输入的三种路线。而将工艺动作过程分解成若干执行动作

时应充分考虑这些执行动作能否用比较简单的执行机构来完成。同时也应考虑前后两执行动作衔接的协调和有效的配合。

总之,分解工艺动作过程要符合下列原则:

- 1) 动作的可实现性:即分解后的动作能被机构实现,因此设计人员应全面掌握各种机构的运动特性。
- 2) 动作实现机构的简单化:即所需实现的动作尽量采用简单机构,有利于机械运动系统易于设计制造。
- 3) 动作的协调性:前后两执行机构产生的动作要相互协调和有效配合,尽量避免两执行机构产生的运动会发生干涉和运动不匹配。

#### (2) 选择合适的执行机构

对于分解工艺动作过程所得的若干执行动作(亦可称工艺动作)要进行详细分析,包括他们的运动规律要求、运动参数、动力性能、正反行程所需曲柄转角等等。

选择执行机构应符合下列要求:

- 1) 执行构件的运动规律与工艺动作的一致性;
- 2) 执行构件的位移、速度、加速度(包括角位移、角速度、角加速度)变化要有利于完成工艺动作;
- 3) 前后执行构件的工作节拍要基本一致,否则无法协调工作;
- 4) 执行机构的动力特性和负载能力要能胜任工艺动作要求;
- 5) 机械运动系统内各执行机构应满足相容性,即输入输出轴相容、运动相容、动力相容、精度相容等等;
- 6) 机械运动系统尺寸紧凑性,即要求各执行机构尺寸尽量紧凑。有利于机械整体尺寸缩小。

#### (3) 网络计划技术法进行执行机构系统的集成

1) 网络图的四要素 网络图是网络计划技术的基础。利用网络图可以对执行机构系统工作进行集成,表达各执行机构工作在时间轴上的先后顺序和相互关系。网络图由箭线和结点组成,它是有向、有序的网络图。网络图可简称为网络。网络图中有四个基本要素:工作、结点、箭线和线路。

(a) 工作:系统中消耗时间和能量的活动称为工作或工序。工作从开始到完成的时间称为延续时间。相对于某一工作,在其前的工作称紧前工作,在其后的工作称紧后工作。

(b) 结点:在网络图中用结点代表工作,用圆圈表示,并且编上数码,填入圆圈中。网络图中的第一个结点,它表示工作的开始,称为起始结点。最后一个结点标志着这项工作的完成,成为终止结点。其他结点称为中间结点。

(c) 箭线:在网络图中用箭线表示工作之间的工艺关系和时序关系。

(d) 线路:网络图中从起始结点出发,沿着箭线方向到达终止结点的通道上连续通过的一系列箭线与结点,最后到达终止结点所经过的通路称为线路。线路上所有工作持续时间之和称为线路长度。网络图中最长的线路称为关键线路。关键线路的长度称为网络的计算工期,对于机械运动系统就是机械工作周期。

#### 2) 网络图的绘制 网络图的绘制步骤如下:

(a) 确定工作项目:对于机械运动系统就是确定有一定程序的执行动作,把各执行机构可以看作机械运动系统的子系统。

(b) 确定工作之间的关系:对于机械运动系统就是确定执行动作的工艺关系和时序关系,即确定该执行动作的紧前工作(动作)和紧后工作(动作)是哪些。

(c) 确定工作的延续时间:时间单位在机械运动系统来说是分、秒,也可以是主轴的角度。

(d) 列出工作明细表:包括工作代号、工作名称、紧前工作、紧后工作、延续时间。

(e) 绘制网络图:此时应遵守以下规则:

- (I) 结点之间的关系必须按工作之间关系绘制;
- (II) 网络图中不能出现回路,即从某一结点出发,顺箭线方向不能回到该结点的线路;
- (III) 在两个相邻结点之间只有一条箭线;
- (IV) 网络图中有而且仅有一个起始结点和终止结点;
- (V) 避免使用反向箭线。

#### 3) 网络图的时间参数计算

##### (a) 工作的最早时间

工作的最早开始时间,最早完成时间是指工作在可能条件下最早开始和结束的时刻。若工作*i*其紧前工作为*h*、紧后工作为*j*,工作*i*延续时间为 $D_i$ 。而其最早开始时间为 $T_i^{ES}$ 。则最早完成时间 $T_i^{EF}$ 为

$$T_i^{EF} = T_i^{ES} + D_i$$

##### (b) 工作的最迟时间

若最迟开始时间为 $T_i^{LS}$ ,其最迟结束时间 $T_i^{LF}$ 为

$$T_i^{LF} = T_i^{LS} + D_i$$

##### (c) 工作的时差、关键工作和关键路线

工作*i*的自由时差记为 $F_i^F$ ,其计算式为

$$F_i^F = T_j^{ES} - T_i^{ES} - D_i$$

或

$$F_i^F = T_j^{EF} - T_i^{EF}$$

工作的总时差是指不影响工期的前提下,该工作所拥有的机动时间的极限值。工作总时差 $F_i^T$ 为

$$F_i^T = T_i^{LS} - T_i^{ES}$$

或

$$F_i^T = T_i^{LF} - T_i^{EF}$$

若 $F_i^T = 0$ 或 $F_i^T$ 为最小值,则该工作*i*为关键工作。

要调整机械运动系统的工作周期,须调整关键路线上的工作时间。

4) 搭接网络计划技术 为了缩短机械运动系统的工作周期,可以采用紧后工作与紧前工作平行施工,工作间的这种关系称之为搭接关系。

5) 举例 如图 47.4-2 所示的纸盒成型及充填包装的工艺动作过程,它可以分解为 9 个工艺动作,他的执行工作的网络图如图 47.4-13 所示。

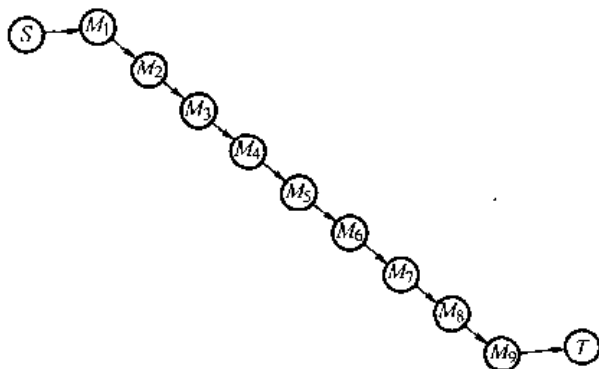


图 47.4-13 纸盒成型及填充包装工艺动作过程网络图

其中  $M_1$ ——纸盒料坯送出、 $M_2$ ——纸盒初次成型、 $M_3$ ——纸盒一端盒口闭合、 $M_4$ ——纸盒一端盒口封口、 $M_5$ ——纸盒翻转  $90^\circ$ 、 $M_6$ ——纸盒送进、 $M_7$ ——物料充填纸盒、 $M_8$ ——另一端盒口闭合、 $M_9$ ——另一端盒口最后封口。

图 47.4-13 反映了 9 个工艺动作之间关系的网络图,两相邻结点之间的箭线表示两个相邻动作的工艺关系和时序关系。

分析图 47.4-2 工艺动作过程分解后的动作情况,我们可以选定机构  $M_1$  完成纸盒料坯从储存器中送出和纸盒成型两个动作,一般可选择实现直线运动的机构;选择机构  $M_3$  完成盒口闭合动作,一般可选择左右先后摆动机构;选择机构  $M_4$  完成盒口封口的动作,一般可选择直线移动的机构;选择机构  $M_5$  完成纸盒翻转  $90^\circ$  的动作,一般可选择刚体导引机构;选择机构  $M_7$  纸盒送紧及物料充填纸盒的动作,用直线运动来完成;选择机构  $M_8$  完成盒口闭合动作;选择机构  $M_9$  完成盒口最后封口的动作。

由此可见纸盒成型的工艺动作过程的 9 个工艺动作作用 7 个机构—— $M_1$ 、 $M_3$ 、 $M_4$ 、 $M_5$ 、 $M_7$ 、 $M_8$ 、 $M_9$  来完成。

根据图 47.4-14 执行动作网络图可以定出每个机构的最大回程和工作行程的时间:机构  $M_1$  为  $T_1$ 、 $T_2$ ,机构  $M_3$  为  $T_2$ 、 $T_3$ ,机构  $M_4$  为  $T_3$ 、 $T_4$ ,机构  $M_5$  为  $T_4$ 、 $T_5$ ,机构  $M_7$  为  $T_5$ 、 $T_6$ ,机构  $M_8$  为  $T_6$ 、 $T_7$ ,机构  $M_9$  为  $T_7$ 、 $T_8$  等。

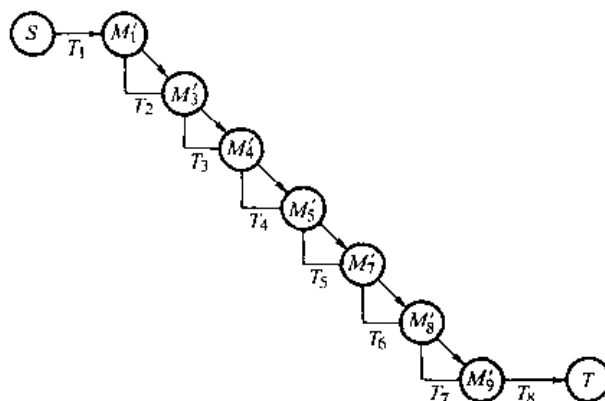


图 47.4-14 纸盒成型及填充包装机执行机构执行动作网络图

### 3 执行机构的协调设计

根据机械的工作原理和工艺动作分析所设计、构思的机械工艺路线方案是机械运动原理方案设计的重要依据。由机械工艺路线,可以进行执行机构的选定和布局。此时必须考虑机械中的各执行机构的协调设计。为此,必须深入了解机械运动方案所采用各执行机构和传动系统的类型、工作原理和运动特点,同时要了解执行机构协调设计的目的和要求,掌握有关协调设计的基本方法。也就是说,要采用集成的设计方法。

#### 3.1 机器的机构传动系统类型和工作原理

机器的机构传动系统是机器的重要组成部分,它决定了机器的执行构件实现工作行程和空行程的方式,影响各执行构件的协调运动和机器的生产率。在机器的机构传动系统的设计中,要求解决机械程序控制和实现空程的方式这两个比较关键的问题,使各执行构件按一定顺序动作并在保证较高生产率的前提下实现空行程。

##### (1) 机械程序控制的基本形式

为了使机器的整个工作循环中各个工序动作严格地按一定顺序进行,每台机器都有专门的程序控制系统。机器的程序控制系统主要有两种类型:一是机械控制方式,二是电子控制方式。目前机器中常用的控制方式有以下几种形式:

1) 分配凸轮轴方式 即在分配轴上安装多个凸轮,每个凸轮是一个执行构件的驱动构件。因此,分配轴每转一周所需要的时间就是产品加工工作循环所需时间,因此,分配轴每转一周就可完成一个产品加工的工艺动作过程。

2) 辅助凸轮轴方式 辅助凸轮轴是机器中用来实现机器部分空行程动作的机械控制方式。辅助凸轮轴是做周期性的间歇旋转,即在需要实现这部分空行程

时才旋转。它的间歇旋转是由分配凸轮轴来控制的。例如, 机器中的送料、夹料等机构的空行程实现就是采用辅助凸轮轴。

3) 曲柄轴方式 利用曲柄的错位使各执行机构按一定顺序来动作, 对于不能将曲柄布置在一根轴上的机器, 还应采用机械传动方式, 将各曲柄的转动予以同步。

4) 机电结合的程序控制方式 在自动机械中还有采用分配凸轮轴上的信号凸轮来控制电路的接通与关闭, 以及通停电时间的长短, 以此来控制各执行机构的动作。

5) 电子控制方式 目前应用最广泛的是用可编程控制器(PLC)来进行控制。它是一种在工业环境中使用的数字操作的电子系统, 它在可编程存储器内部储存用户设计的指令, 这些指令用来实现特殊的功能, 诸如逻辑运算、顺序操作、定时、计数、以及算术运算和通过数字或模拟输入、输出来控制各种类型的机械或过程, 以实现生产自动化。其特点是: ①控制程序可变, 具有很好的柔性; ②具有高度可靠性; ③功能完善; ④易于掌握、便于维修; ⑤体积小、省电; ⑥价格低廉。

总之, 机械程序控制的具体实施, 还应根据具体情况来确定。

(2) 机器实现空程的方式

根据机器中实现空程的方式, 机器可分为以下 3 种类型:

1) 空程转角固定的机器 关于空程转角固定的机器, 它的分配轴转 1 圈完成 1 个工作件的加工, 即一个工艺动作过程, 参见图 47.4-15。图 47.4-15b 表示安装在分配轴上的凸轮廓线的工作行程和空行程的转角组成情况。凸轮廓线是由工作行程部分和空行程部分所组成。其中  $\alpha$  表示分配轴完成工作行程所转的角度,  $\beta$  表示分配轴完成空行程所转的角度。

这类机器的生产率  $Q$  为:

$$Q = \frac{n_{\pm}}{N_0} \quad (47.4-5)$$

式中  $n_{\pm}$  —— 主轴转速 (r/min);

$N_0$  —— 加工一个产品所需的主轴转数。

若  $N_1$  表示完成工作行程所需主轴的转数:

$$N_0 = \frac{2\pi N_1}{\alpha}$$

式 (47.4-5) 可写成:

$$Q = \frac{n_{\pm}}{N_1} \cdot \frac{\alpha}{2\pi}$$

由于连续生产率  $K = \frac{n_{\pm}}{N_1}$ ,  $\alpha = 2\pi - \beta$ , 所以:

$$Q = K \cdot \frac{\alpha}{2\pi} = K \left( 1 - \frac{\beta}{2\pi} \right) \quad (47.4-6)$$

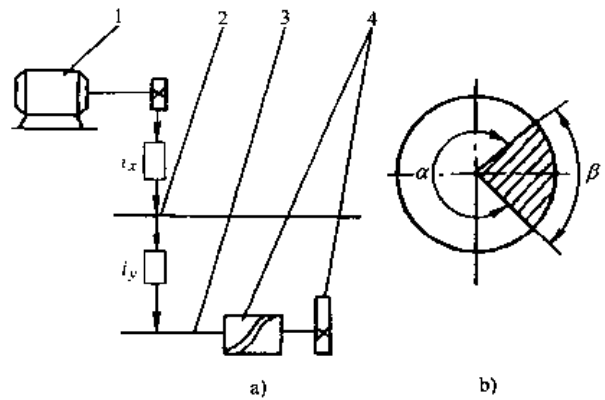


图 47.4-15 空行程转角固定的机器

1—电动机 2—主传动轴 3—分配轴 4—凸轮

$i_x$ —主传动轴转速调整环节  $i_y$ —分配轴转速调整环节

由此得出机器的生产率系数  $\eta$  为:

$$\eta = 1 - \frac{\beta}{2\pi} \quad (47.4-7)$$

根据式 (47.4-6) 可以得出:

(a) 由于  $\beta = \text{常数}$ , 所以机器生产率  $Q$  将随连续生产率  $K$  正比例地增加;

(b) 当机器连续生产率  $K$  很小时, 即完成工作行程的时间很长, 亦即机器的工作循环时间  $T$  很大。由于  $\beta$  一定, 必定使完成空行程的时间很长, 造成不必要的时间损失。这就需要限制工作循环时间  $T$  的数值,  $T$  的极限值表示为  $T_{\max}$ ;

(c) 当机器连续生产率  $K$  很大时, 即完成工作循环的时间  $T$  极短, 也就是分配轴转速极快。由于  $\beta$  一定, 使完成空程的时间很短, 易产生不能可靠地工作。这就需要延长工作循环时间  $T$ , 它的极限值为  $T_{\min}$ 。

综上所述, 这类机器的生产率只适应于一定范围, 即  $T_{\min} < T < T_{\max}$ 。这类机器的机构传动系统由于空行程

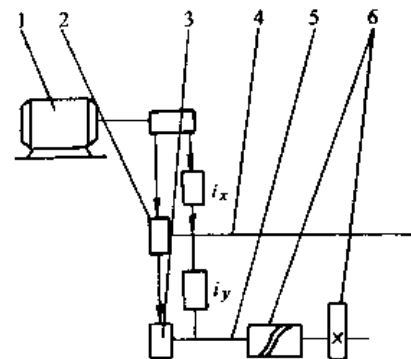


图 47.4-16 空程时间固定的机器

1—电动机 2—快速接合器

3—超越离合器 4—主传动轴

5—分配轴 6—凸轮

$i_x, i_y$ —转速调整环节

转角固定而比较简单。

2) 空行程时间固定的机器 图 47.4-16 为关于空行程时间固定的机器,它的分配轴具有两条传动路线,一是通过变速机构使分配轴作慢速转动以完成工作行程;二是通过快速离合器、超越离合器,直接使分配轴作快速转动以完成空行程。离合器的接合与否,由分配轴上的凸轮机构来控制。这种机器的主要特点就是不管工件加工时间的长短如何,它的空行程时间保持不变。因此,这类机器的传动系统比较复杂,要增加离合器、超越离合器和控制凸轮等。它适合于加工工作循环时间  $T$  较大的产品。在传动系统中加超越离合器的作用是为了防止分配轴发生运动干扰。

### 3.2 机器执行机构的协调设计

设计机器时,在选定了它的各个执行机构之后,下一步就是根据生产工艺路线方案,使这些执行机构进行合理布局 and 相互协调动作,以确保进行正常的产品加工。否则,会使机器不能进行正常工作,严重的还会损坏机件和被加工的工件(产品),造成事故。

机器执行机构的协调设计应满足以下的要求:

- (1) 机器各执行机构的动作过程和先后次序要符合机器的生产工艺路线方案所提出的要求,否则就无法满足机器的生产工艺,也就不能实现机器的工作原理。
- (2) 机器各执行机构的运动循环的时间同步化,亦即各执行机构的运动循环时间间隔相同或按生产工艺过程要求成一定的倍数。使各执行机构的动作不但保

证在时间上有顺序关系,而且能够实现周而复始的循环协调动作。

(3) 机器各执行机构在运动过程中,不仅要在时间上保证一定的顺序关系,而且在一个运动循环的时间间隔内,运动轨迹不相互干涉。同时,为了保证机器的工作质量,既不能使动作先后顺序的间隔时间太长,又不能使动作先后顺序间隔时间太短。这称为机器各执行机构运动循环空间同步化。动作先后顺序的间隔时间太长,会使机器生产率下降;动作先后顺序间隔太短,有可能使执行构件产生相互干扰。

为了说明协调设计的目的和要求以及协调设计的方法,我们举粉料压片机为例来加以介绍(参阅图 47.4-17)。粉料压片机的机械运动简图如图 47.4-17c 所示,它由上冲头(六杆肘杆机构)、下冲头(双凸轮机构)、料筛传送机构(凸轮连杆机构)所组成。料筛由传送机构把它送至上、下冲头之间,通过上、下冲头加压把粉料压成有一定紧密度的药片。根据生产工艺路线方案,此粉料压片机必须实现以下 5 个动作(见图 47.4-17a):

- ① 移动料斗至模具的型腔上方准备将粉料装入型腔,同时将已经成型的药片推出;
- ② 料斗振动,将料斗内粉料筛入型腔;
- ③ 下冲头下沉至一定深度,以防止上冲头向下压制时将型腔内粉料扑出;
- ④ 上冲头向下,下冲头向上,将粉料加压并保压一定时间,使药片成型较好;
- ⑤ 上冲头快速退出,下冲头附着将成型工件(药片)推出型腔,完成压片工艺过程。

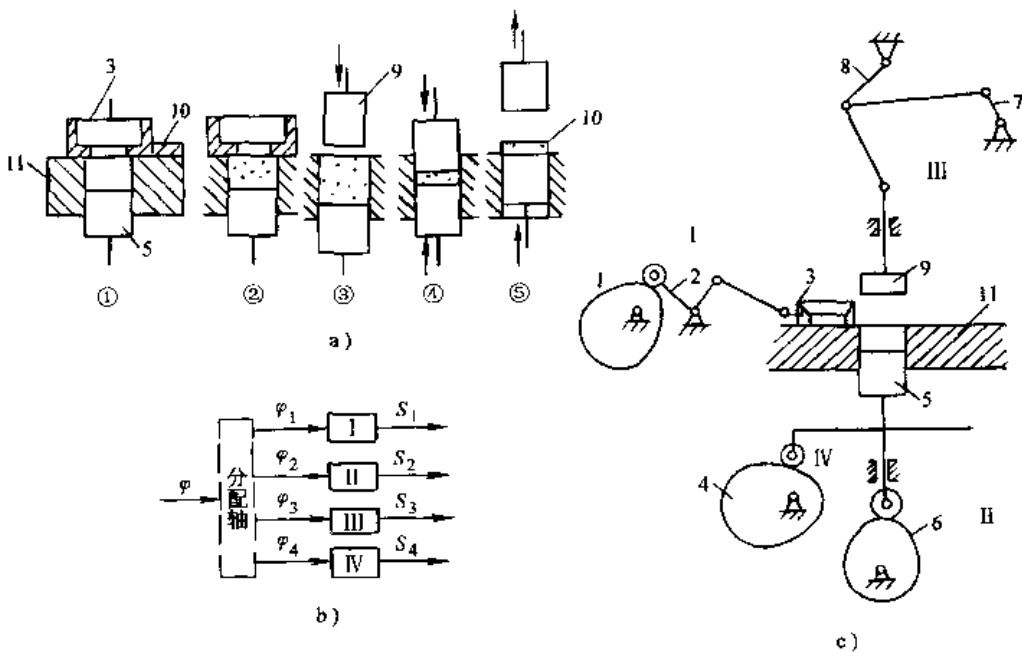


图 47.4-17 粉料压片机

1、4、6—凸轮 2、7、8—连杆 3—料斗 5—下冲头 9—上冲头 10—药片 11—模具

这 5 个动作如图 47.4-17a 所示, 机器各执行机构的动作过程和先后次序就要按此进行, 否则无法实现机械的粉料压片工艺。从图 47.4-17c 所示机器的机械运动简图可见, 它由 4 个执行机构来完成上述 5 个动作。凸轮连杆机构 I 完成工艺动作①、②, 凸轮机构 II 完成动作③; 平面多杆机构 III 及凸轮机构 IV 协调配合完成动作④、⑤。

粉料压片机执行机构的运动协调设计可从以下两方面来阐述。

#### (1) 各执行机构的动作在时间上协调配合

在此粉料压片机中执行构件为 3、9、5、4 个执行机构的原动件为 1、4、6、7。为了使各执行机构的运动循环的时间同步化, 我们可以将原动件 1、4、6、7 安装在同一根分配轴上, 或用一些传动机构把它们连接起来, 以实现原动件转速相同和相互间有一定的相位差。在同一根分配轴上的构件 1、4、6、7 只要按动作顺序安排, 就可实现周而复始的循环协调动作。如果采用一些传动机构把构件 1、4、6、7 连接起来, 应使它们在同一转速下运转并保持动作顺序。在某些其他类型机器中也可使它们不在同一转速下运转, 此时各原动件转速比值为某一整数比, 以实现周而复始的协调动作。一般以原动件最低转速所对应的运动循环为整机的运动循环, 较高转速的构件一般应作间歇运动, 以实现各执行构件动作的协调配合。

#### (2) 各执行机构的动作在空间上协调配合

在粉料压片机中, 执行构件 3、9 的两个运动轨迹是相交的, 故在安排两执行构件的运动时, 不仅要注意到时间上的协调, 还要注意到空间位置上的协调——空间同步化。亦即使两执行构件在运动空间内不相互干扰。时间协调与空间同步化有密切关系。

### 3.3 执行机构协调设计的分析计算

#### (1) 各执行机构运动循环时间同步化的计算

1) 机器最大工作循环周期  $T_{\max}$  的确定 机器最大工作循环周期  $T_{\max}$  是机器各执行机构工作必须的工作循环时间之和。例如, 图 47.4-17c 所示的粉料压片机, 它由 4 个执行机构所组成, 各个执行机构的动作过程一般可以分解为: 工作行程—最远停歇—回程(空行程)—最近停歇这 4 个阶段, 它们各自的工作循环时间为  $T_{p1}$ 、 $T_{p2}$ 、 $T_{p3}$ 、 $T_{p4}$ , 则机器最大工作循环周期  $T_{\max} = T_{p1} + T_{p2} + T_{p3} + T_{p4}$ 。这样的计算结果虽然能确保各执行机构的动作次序和时间要求, 但是很显然这并不合理。

2) 机器最小工作循环周期的确定 机器最小工作循环周期  $T_{\min}$  是机器各执行机构工作循环时间的最大值  $T_{p\max}$ 。

例如, 粉料压片机, 它的 4 个执行机构中机器 I 的工作循环时间  $T_{p1}$  为最大, 则  $T_{\min} = T_{p1}$ 。如果以  $T_{\min}$  作为机器的工作循环周期, 这样也不一定合理, 要在  $T_{\min}$  时间内同时完成 4 个执行机构的动作, 又不能在空间内产生干涉, 有时往往是比较困难的, 因此必须增大工作循环周期。只有当机器的执行机构较少而且相互间时间协调较好时, 经过努力才可以在  $T_{\min}$  时间内完成工艺动作次序和要求。

3) 确定合理的机器的工作循环周期  $T$  在确保各执行构件工作行程、回程的要求下, 采用各执行机构的工作循环部分重合的方法来合理确定机器的工作循环周期  $T$ , 以实现工艺过程所需的动作次序和相应的时间要求。为了尽量缩短机械工作循环周期  $T$ , 提高机器的生产率, 一般可以采取两个措施: 一是将各执行机构的工作行程和回程时间在可能条件下尽量缩短; 二是在前一个执行机构的回程结束之前, 后一执行机构就开始作工作行程, 这就是利用两执行构件空间裕量, 在不产生相互干涉的条件下, 采用“偷时间”的办法。对于具有较多执行机构的机器采用上述两方面措施之后, 其效果是十分明显的。

4) 确定各执行机构分配轴的转速和对应起始角 大多数机器各执行机构的原动件都安装在同一分配轴上, 我们可以根据机器工作循环的周期  $T$  算出分配轴的转速  $n_{分}$ :

$$n_{分} = \frac{60}{T} \quad (47.4-8)$$

其中,  $T$  的单位为 s。

一般在机器中, 将某一执行机构工作行程起始点作为零位, 根据机器各执行机构的生产工艺动作次序的安排不难求出各执行机构工作行程的起始角。

#### (2) 各执行机构运动循环空间同步化的计算

图 47.4-18 所示为饼干包装机的两个折侧边的执行机构。M 点是两折边器运动轨迹的交点, 说明空间同步化设计不好将会产生空间上的相互干涉。如果两折边先后顺序动作的间隔时间太长, 虽然空间上相互不干涉, 但会使折过边的包装纸重新弹回虚线位置, 无法保证包装质量。反之两折边机构先后顺序动作的时间间隔太短, 会使两折边器在空间上相碰, 使机件损坏。

空间同步化计算前需已知执行机构的动作顺序, 各执行机构执行构件的实际位移曲线图, 各执行机构的机构简图。从而合理确定各执行机构的运动错位量  $\Delta t$ 。图 47.4-19 表示图 47.4-18 的饼干包装机左右折边机构的位移曲线图  $\varphi-t$ 。交点 M 所对应的角位移  $\varphi_M^L$  和  $\varphi_M^R$ , 它们在  $\varphi-t$  曲线上的点  $M_1$ 、 $M_2$ 。由于存在错位量  $\Delta t$ , 使左右两折边机构不会产生相碰。如果  $M_1$ 、 $M_2$

重合于一点,则肯定会产生执行构件相碰的现象,解决办法是将左右两折边机构的执行构件的位移曲线加以改变。如用时间上的错位等。

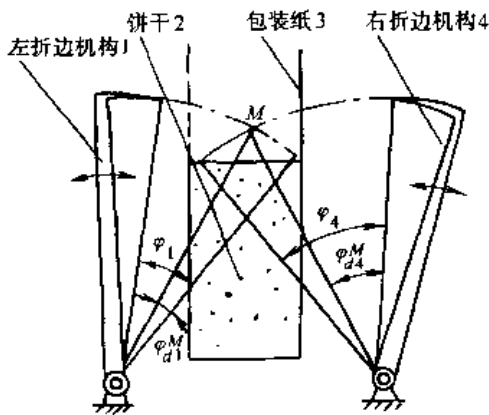


图 47.4-18 饼干包装机的两个折侧边机构

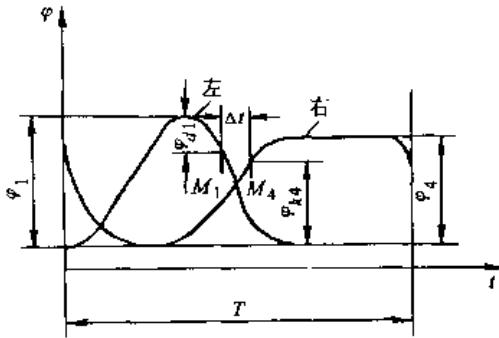


图 47.4-19 饼干包装机折边机构的位移曲线

## 4 机械运动循环图设计

### 4.1 机器的运动分类

根据机器所完成功能及其生产工艺的不同,它们的运动可分为无周期性循环和有周期性循环两大类:一类为无周期性循环的机器如起重运输机械、建筑机械、工程机械等。这类机器,它们的工作往往没有固定的周期性循环,随着机器工作地点、条件的不同而随时改变;另一类为有周期性循环的机器,如包装机械、轻工自动机、自动机床等。这类机器中的各执行构件,每经过一定的时间间隔,它的位移、速度和加速度便重复一次,完成一个运动循环。在生产中大部分机器都属这类具有固定运动循环的机器。它们是我们这节中要讨论的对象。

为了保证具有固定运动循环周期的机械完成工艺动作过程时,各执行构件间的动作的协调配合关系,在设计机械时,应编制出用以表明在机械的一个运动循环中,各执行构件运动配合关系的机械运动循环图(也叫机器工作循环图)。在编制机械运动循环图时,必须

从机械的许多执行构件(或输入构件)中选择一个构件作为运动循环图的定标件,用它的运动位置(转角或位移)作为确定各个执行构件的运动先后次序的基准,表达机械整个工艺动作过程的时序关系。

### 4.2 机械的运动循环周期

机械的运动循环是指一个产品在加工过程中的整个工艺动作过程(包括工作行程、空回行程和停歇阶段)所需要的总时间,它通常以  $T$  表示。在机械的工作循环内,其各执行机构必须实现符合工件(产品)的工艺动作要求和确定的运动规律、有一定顺序的协调动作。

执行机构完成某道工序的工作行程、空回行程(回程)和停歇所需时间的总和,称为执行机构的运动循环周期。各执行机构的运动循环与机器的工作循环,一般来说在时间上应是相邻的。但是,也有不少机器,从实现某一工艺动作过程要求出发,某些执行机构的运动循环周期与机器的工作循环周期并不相等。此时,机器的一个工作循环内有些执行机构可完成若干个运动循环。

执行机构的运动循环周期  $T_p$  通常由三部分组成:

$$T_p = t_{工作} + t_{回程} + t_{停歇}$$

式中  $t_{工作}$  —— 执行构件工作行程时间;  
 $t_{回程}$  —— 执行构件空回行程时间;  
 $t_{停歇}$  —— 执行构件停歇时间。

### 4.3 机器的工作循环图

机器的工作循环图是表示机器各执行机构的运动循环在机器工作循环内相互关系的示意图,它也可称为机器的运动循环图。机器的生产工艺动作顺序是通过拟定机器工作循环图选用各执行机构来实现的。因此,工作循环图是设计机器的控制系统和进行机器调试的依据。

#### (1) 执行机构的运动循环图

表示执行构件的一个动作过程(包括工作行程、空回行程和间歇停歇停顿阶段),称为执行机构的运动循环图。

图 47.4-20 所示的自动压痕机,其压痕冲头的上下运动是通过凸轮来实现的。冲头的运动循环由三部分组成:冲压行程所需时间  $t_s$ , 压痕冲头的保压停留时间  $t_0$  以及回程所需时间  $t_d$ 。因此压痕冲头一个循环所需时间  $T_p$  为:

$$T_p = t_s + t_0 + t_d \quad (47.4-9)$$

用图形表示执行构件运动循环的方式通常有 3 种:

1) 直线式运动循环图 以一定比例的直线段表示

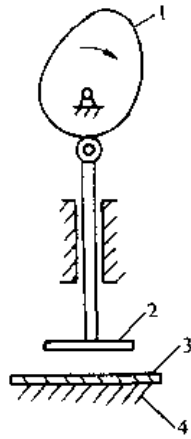


图 47.4-20 自动压痕机的最简结构型式

- 1—凸轮 2—压痕冲头
- 3—压印件 4—下压痕模

运动循环各运动区段的时间(图 47.4-21a)。

这种表示方法最简单,但直观性很差(例如压痕冲头在每一瞬时的位置无法从图上看),且不能清楚地表示与其他机构动作间的相互关系。

2)圆形运动循环图 将运动循环的各运动区段的时间及顺序按比例绘于圆形坐标上(图 47.4-21b)。

此法直观性强,尤其对于分配轴每转一周为一个机械工作循环者,有很多方便之处。但是,当执行机构太多时,需将所有执行机构的运动循环图分别用不同

直径的同心圆环来表示,则看起来不很方便。

3)直角坐标运动循环图 以直角坐标表示各执行构件的各个运动区段的运动顺序及时间比例,同时还表示出执行构件的运动状态(图 47.4-21c)。

此法直观性最强,比上述两种运动循环图更能反映执行机构运动循环的运动特征。所以在设计机器的工作循环图时,最好采用直角坐标运动循环图。

(2) 机器的工作循环图

机器的工作循环图是机器中各执行机构的运动循环图按同一时间(即按某一转轴的转角)比例绘制并组合起来的总图。该图应以某一主要执行机构的起点为基准,表示其余各执行机构的运动循环相对于该主要执行机构的动作顺序。

图 47.4-20 所示的自动压痕机最简单的结构型式是由压痕机构和送料机构所组成。如果要考虑成品自动落料,还应有一个落料机构。在图中送料机构没有表示出来,送料机构的运动循环周期  $T_p'$  为:

$$T_p' = t_k' + t_o' + t_d'$$

- 式中  $t_k'$ ——送料机构的上料所需时间;
- $t_o'$ ——送料到位后执行机构的停歇时间;
- $t_d'$ ——送料机构的回程所需时间。

很显然,送料机构的运动循环周期  $T_p'$  应与压痕机构的运动循环周期  $T_p$  相等。

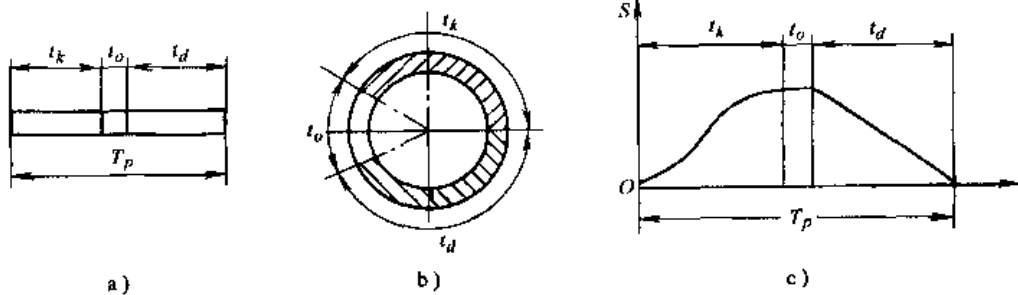


图 47.4-21 执行构件的运动循环图

绘制压痕机的工作循环图,可以将压痕冲头的最高点作为起点,并以它作为基准画出此两执行机构的运动循环图,它们组合在一起就成为压痕机的工作循环图,如图 47.4-22 所示。它是按直角坐标法画出的运动循环图,工作行程由起点开始向上表示,空回行程由最远点回至起点表示,这与实际执行构件的上下、左右运动无直接关系。用直角坐标表示的运动循环图还可以表示出工作行程和空回行程中执行构件的运动规律。

送料机构的运动循环的动作必须与压痕冲头的运动循环的动作相协调,即在压痕冲头作向下冲压运动时,送料机构应停歇不动,当压痕冲头作回退运动和停

歇时,送料机构可作上料动作。在具体制定它们的运动循环图时,只要动作协调、互不干涉,可以进行小范围的调整。

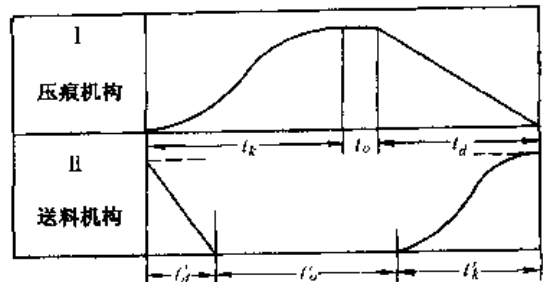


图 47.4-22 压痕机工作循环图



#### 4.4 拟定机器工作循环图的步骤和方法

##### (1) 拟定机器工作循环图的步骤

1) 分析加工工艺对执行构件的运动要求(如行程或转角的大小,对运动过程的速度、加速度变化的要求等)以及执行构件相互之间的动作配合要求。

2) 确定执行构件的运动规律,这主要是指执行构件的工作行程、回程、停歇等与时间或主轴转角的对应关系,同时还应根据加工工艺要求确定各执行构件工作行程和空回行程的运动规律。

3) 按上述条件绘制机器工作循环草图。

4) 在完成执行机构选型和机构尺度综合后,再修改机器的工作循环图。具体来说,就是修改各执行机构的工作行程、空回行程和停歇时间等的大小、起始位置以及相对应的运动规律。

根据初步拟定的执行构件运动规律设计出的执行机构,常常由于布局和结构等方面的原因,使执行机构所实现的运动规律与原方案不完全相同,此时就应根据执行构件的实际运动规律修改机器工作循环草图。如果执行机构所能实现的运动规律与工艺要求相差很大,这就表明此执行机构的选型和尺寸参数设计不合理,必须考虑重新进行机构选型或执行机构尺寸参数设计。

5) 拟定自动控制系统,控制元件的信号发出时间及其工作状态,并将它们在机器工作循环图上表示出来,得到完整的机器工作循环图。

##### (2) 机器工作循环图的设计要点

1) 以工艺过程开始点作为机器工作循环的起始点,并确定开始工作的那个执行机构在工作循环图上的机构运动循环图,其他执行机构则按工艺动作顺序先后列出。

2) 不在分配轴上的凸轮,应将其动作所对应的中心角,换算成分配轴相应的转角。

3) 尽量使各执行机构的动作重合,以便缩短机器工作循环的周期,提高生产率。

4) 按顺序先后进行工作的执行构件,要求它们前一执行构件的工作行程结束之时,与后一执行构件的工作行程开始之时,应有一定的时间间隔和空间裕量,以防止两机构在动作衔接处发生干涉。

5) 在不影响工艺动作要求和生产率的前提下,应尽可能使各执行机构工作行程所对应的中心角增大些,以便减小凸轮的压力角。

#### 4.5 机器工作循环图的作用

1) 保证执行构件的动作能够紧密配合,互相协调,使机器的工艺动作过程顺利实现;

2) 为计算和研究、提高机器生产率提供了依据;

3) 为下一步具体设计各执行机构提供了初始数据;

4) 为装配、调试机器提供依据。

综上所述,拟定机器工作循环图是机器设计过程中一个重要的设计内容,它是提高机器设计的合理性、可靠性和生产率必不可少的工作。

#### 4.6 机械运动循环图设计举例

下面我们以三面切书自动机的运动循环图设计为例来说明机械运动循环图的使用。

##### (1) 三面切书自动机的工艺示意图

如图 47.4-23 所示,它由送料执行构件、压书执行构件、两侧切书刀执行构件、横切书刀执行构件、书本、工作台等组成,其工艺路线为:将书本用送料机构 I 至切书工位;然后用压书机构 II 将书本压紧;接着用两侧切书刀机构 III 切去前面余边(参阅图 47.4-24)。

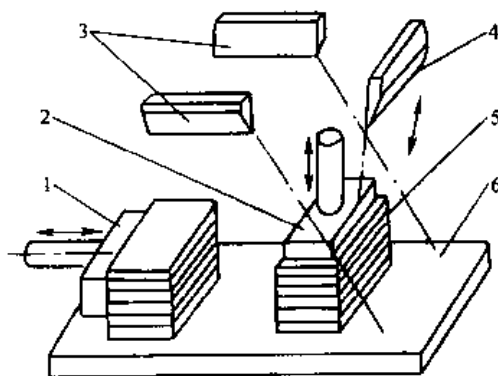


图 47.4-23 切书自动机工艺示意图

1—送料执行机构 2—压书执行构件  
3—两侧切书执行构件 4—横切书刀  
执行构件 5—书本 6—工作台

##### (2) 三面切书自动机运动简图

由图 47.4-24 可知,三面切书自动机的机械运动,是由 4 个执行机构来完成上述工艺动作的。具体说明

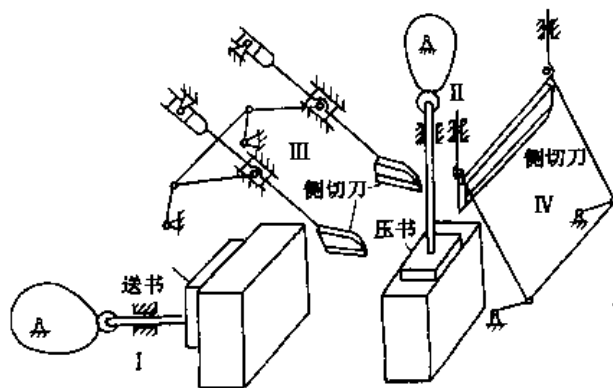


图 47.4-24 切书自动机运动简图

如下：

1) 送料机构 I，它将输送带上输送过来的有一定高度的书本送至切书工位。图中采用了凸轮机构。

2) 压书机构 II，它将在切书工位的书本压紧。图中也采用了凸轮机构。

3) 两侧切书刀机构 III，它将已压好的书的两侧切去余边。这里采用平面连杆机构。

4) 横切书刀机构 IV，它将已切去书的两侧余边的书本再切去前面余边。这里也采用了平面连杆机构。

(3) 三面切书自动机的运动循环图的设计

1) 各执行机构运动循环图的设计 三面切书自动机的送料、压书、两侧切书刀以及横切书刀 4 个执行机构，它们动作的先后顺序均可由分配轴上的凸轮或偏心轮机构来控制。为了方便起见，用分配轴转角来表示各执行机构的运动循环图，如图 47.4-25 所示。其中送料机构由工作行程—回程—初始停歇等 3 个阶段组成；压书机构由工作行程—停歇—回程—初始停歇等 4 个阶段组成；侧刀机构为由初始停歇—工作行程—回程等 3 个阶段组成；横刀机构为由工作行程—回程两个阶段组成。图 47.4-25 只是初步表示这 4 个执行机构的运动循环图，在进行执行机构同步化设计之后，可进一步采取修改设计。

2) 执行机构运动循环同步化设计

(a) 送料机构与压书机构的同步化设计 由于该

自动机裁切书本的高度有一定变化范围 ( $H_{min} \sim H_{max}$ )，因此压书板的行程终点也得相应变化。

图 47.4-26a 表示压书机构，图 47.4-26b 为它的位移曲线。根据被切书本的最大高度  $H_{max}$ ，找出与此对应的曲线上的 A 点，以及 A 点所处的分配轴转角  $\phi_{H_{max}}$  值。因此，送料机构必须在  $\phi_{H_{max}}$  之前将书本送到切书工位，即在分配轴转角为  $(\phi_{H_{max}} - \Delta\phi)$  时完成送料行程。这就是送料机构与压书机的同步化条件。

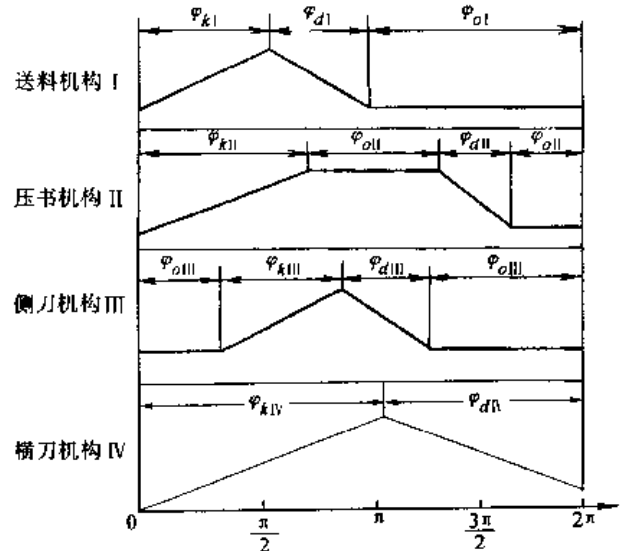


图 47.4-25 切书自动机执行机构运动循环图

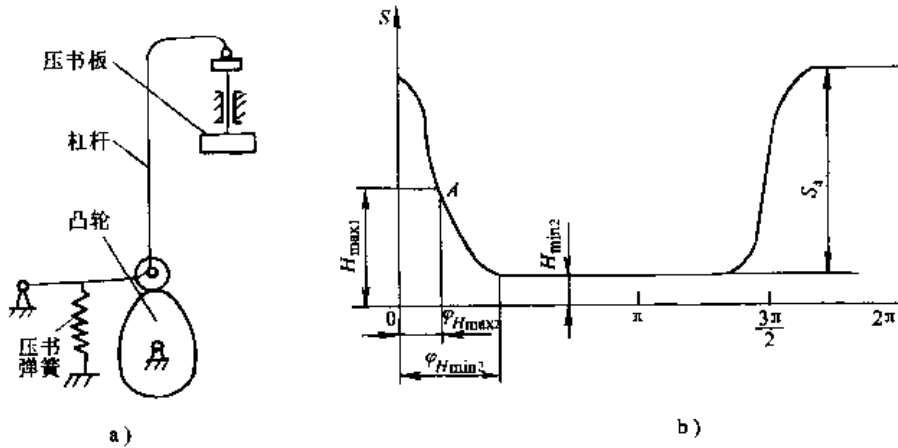


图 47.4-26 切书自动机的压书机构及其位移曲线

(b) 侧刀机构与横刀机构的同步化设计 图 47.4-27 所示为侧刀机构与横刀机构的工作简图，两侧刀的运动轨迹与横刀的运动轨迹在空间相交于 M 点，因此必须进行空间同步化设计以免两者产生干涉。

图 47.4-28 所示为侧刀机构与横刀机构的同步图。交点 M 在侧刀和横刀的位移曲线上对应的为  $M_3$ 、 $M_4$ ，由于有错位量  $\Delta\phi$ ，使实现两机构空间同步化。

(c) 压书机构与侧刀和横刀机构的同步化设计 按切书工艺要求，当压书板压紧书本后侧刀才开始裁切书本两侧余边。当横刀裁切完毕后，压书板才能放松被压书本，并退回原位，准备下一叠未切余边的书本的压紧。考虑被切书本的高度变化，应以最小书本高度  $H_{min}$  作为同步化依据。这样能保证在书本最小高度时将书本压紧，对于书本最大高度时也一定能压紧书本。

另外，当横刀裁切至最低位置并返回至  $H_{min}$  时，压书板才可以松开。

3) 绘制三面切书自动机的运动循环图 根据上

述各执行机构的同步化分析结果，修正图 47.4-25 所示的三面切书自动机的运动循环图，最后绘制出如图 47.4-29 所示的三面切书自动机的机械运动循环图。

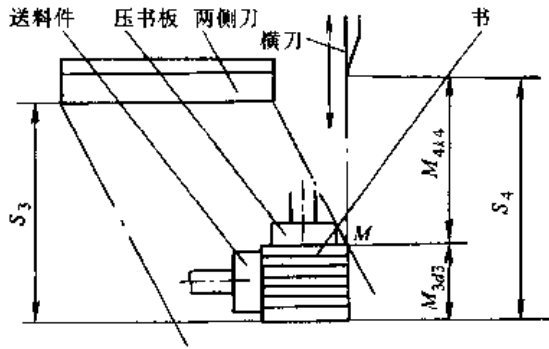


图 47.4-27 切书自动机侧刀与横刀机构工作简图

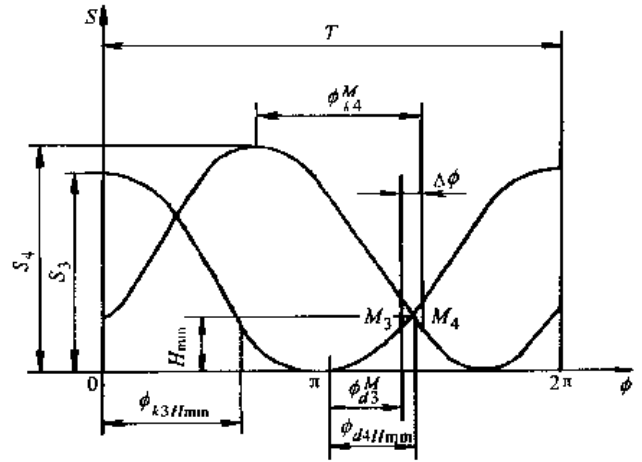


图 47.4-28 切书自动机侧刀与横刀机构同步图

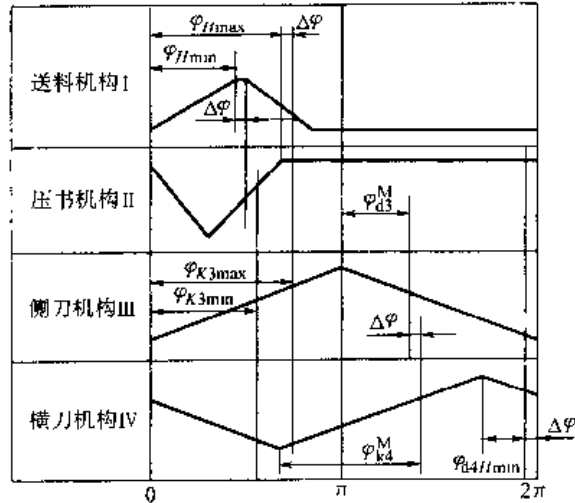


图 47.4-29 修正后的切书自动机机械运动循环图

## 第5章 机械系统运动方案的构思和设计

机械运动系统的方案构思和设计是机械系统设计的重要内容和关键步骤。

### 1 机械系统运动方案设计的主要步骤和内容

#### 1.1 机械系统运动方案设计的主要步骤

机械运动系统方案设计也可以说是机构系统方案的设计。因为机械运动系统是由若干个执行机构组成的。用此机构系统来完成工艺动作过程,达到工作机器实现总功能的要求,如图 47.5-1 所示。

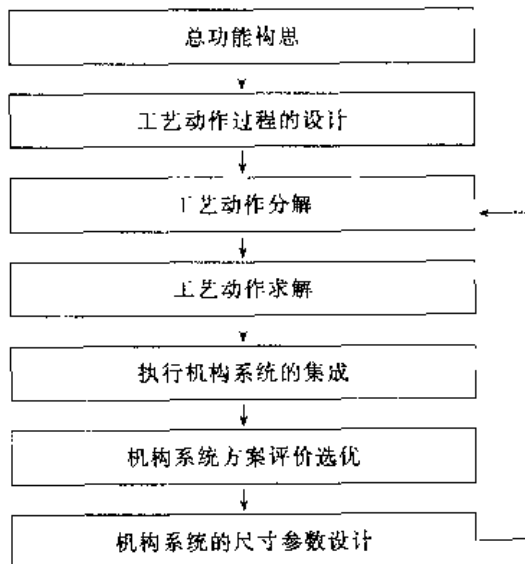


图 47.5-1 机械系统运动方案设计主要步骤  
机械系统运动方案设计的主要步骤如下:

##### (1) 工艺动作过程的构思和设计

工艺动作过程的确定是机械系统运动方案设计的关键步骤,也是体现机械系统运动方案设计的主要特点。

##### (2) 工艺动作分解和求解

工艺动作的分解是指将工艺动作过程分解为若干个执行动作。工艺动作求解是寻求实现分解得到的若干执行动作的可行的机构类型。

(3) 寻求执行机构系统可行方案和进行最优方案的选择

由于实现某一执行动作的机构解不是惟一的,因此执行机构系统可行方案数一般是比较多的,从众多的可行方案中按一定的评价方法来选择综合最优方案。

##### (4) 执行机构系统各机构的尺寸参数设计

执行机构系统方案的确定,一般是先确定各执行机构的类型。机构类型的选定只是说明有可能实现所需的工艺动作过程。为了使机构系统能精确实现机械的工艺动作过程,还应根据机械的运动循环图中所表达的各执行机构的执行构件运动规律,以及各执行构件的运动时间关系来进行各执行机构的尺寸参数设计。如果发现某一执行机构尺寸参数无法实现所需运动规律和运动时间序列,那么就必须改变机构的类型,重新进行机构尺寸参数设计。

#### 1.2 功能原理方案设计

实现同样的总功能,可以有多种不同的工作原理。工作原理可以理解对各种科学原理,包括物理学的、化学的、生物学的等各学科的科学原理。而且在同属物理学内的科学原理也是多种多样的,有力学、光学、声学、热学、电学等等。功能原理就是实现功能的工作原理。一个崭新的功能原理方案就可以创造出一种新颖的机械产品。因此,在构思功能原理方案之前应多熟悉了解各种各样新技术、新工艺、新材料。做到开阔思路、力求创新。要寻求一种新颖的、合理的、最优的功能原理方案是一件相当复杂又十分困难的创新工作,往往需要经过大量的实验验证工作。

下面举例说明功能方案的构思和设计

##### (1) 洗衣机的功能原理方案

洗衣机是为了实现“洁衣”的核心功能,从污物和衣物分离的需要出发,它的功能原理有化学方法的干洗式、力学方法的机械搅拌式以及声学方法的超声波振荡式等。如若采用机械搅拌式的工作原理还可以分为波轮式(漩涡式)、搅拌式(摆动式)和滚筒式(拍洗)等3种方式。如果模拟人手搓动,还可以用搓动式等等。

构思出一种新颖的功能原理,就可以开发出一种新颖的产品。通过这一途径便可使洗衣机的新品种层出不穷。

##### (2) 点火机的功能原理方案

点火机是为了实现“点火”的核心功能,可以利用摩擦取火法得到火柴式点火;利用可燃气体或液体燃烧法得到打火式点火;以及利用电热法得到电热式点火。除了这3种点火方法外,还可以用聚光法点火等等。

##### (3) 加工齿轮的功能原理方案

加工齿轮的功能原理也是多种多样的,要按它的精度要求、制造批量、加工成本、使用场合不同而加以选择。采用铸造法的铸造齿轮;采用冲压法的冲压成型齿轮;采用挤压法的挤压成型齿轮以及切削法的切削齿轮。对于切削法又可分成仿型法(如铣削、拉削)和展成法(如插齿、滚齿)。仿型法加工精度不及展成法。

设计人员在构思和选择机械的功能原理方案时,可以采用发散思维方法,寻求一切可行的功能原理,进行功能原理方案创新设计。大家认识到方案创新不可能是无本之木、无源之水,设计人员应重视在科学原理应用上的新突破或技术基础研究上的新构想。例如采用石英振荡原理获得高精度的计时原理,使钟表功能原理有了新的突破。又如,采用激光技术可以设计出测量船舶尾轴的对中程度。

综合应用已有的科学原理和工程技术也是一种创新。例如,将喷气推进原理同燃气轮机技术相结合,发明了喷气式发动机。又如,利用电磁原理和摩擦原理相结合,发明了电磁离合器。再如,将超声技术同气流原理相结合,发明了超声波洗衣机,等等。

总之,在功能原理方案的设计上多下功夫,多加探索,将会产生意想不到的效果,有可能开发出一种新颖的产品。对于提高机械产品的创新性起着十分重要的作用。功能原理方案对于功能-结构过程的中间部分加以细化,通过功能原理将功能具体化,有利于求得功能载体(或称结构)。对于机械产品设计来说,功能原理又可称之为工作原理,用功能原理可以构思出机械产品的工艺动作过程。

### 1.3 机械系统运动方案设计

机器的功能原理方案的构思和设计,只是提出了实施机器的各分功能的原理方案。对于机械产品来说,从功能原理方案到提供生产用的构形设计图纸,其间还要做不少工作。其中第一步还是要进行机械运动系统方案设计,也就是机构系统方案的设计。具体来说是将功能原理方案所需实施的各分功能,构想出相应的执行动作。此一系列执行动作按运动循环图的顺序构成了机器的工艺动作过程。对各执行动作选择合适的执行机构来加以实现,这些执行机构所组成的机构系统就可实现所需的机器工艺动作过程。图47.5-2表示机械运动系统方案(机构系统方案)设计流程图。由于同一执行动作可以用多个执行机构来实现,因此机构系统方案可以有好几个,通过选择可以得到综合最优的方案。

机械运动系统方案设计时应充分考虑对机器的设计要求:机器所需功率、生产能力、空间尺寸限制、物

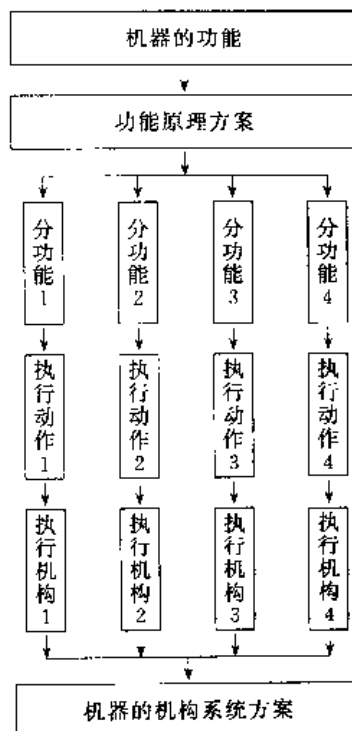


图 47.5-2 机构系统方案设计流程图

料流动方向、工作环境等等。如何满足这些设计 requirements 是机械运动系统方案设计时需一一加以考虑的。

为了得到性能优良、结构简单、工作可靠的机械运动系统方案,设计时应注意如下的几个原则:

#### (1) 在工艺动作过程分解时应注意巧妙

对机器的工艺动作过程分解后所得到的执行动作应该能被常用的机构加以实现。但是利用一些布置巧妙的挡块,可以用一个执行动作完成几个工艺动作。例如,图47.5-3所示方糖折叠式包装,利用一个使方糖产生移动的执行机构,再加上上下两个挡块,使包装纸一次完成上、下、前面三个面的包装。使机械运动系统方案大为简化。如果上、下、前面的包装分别进行,至少要加上两个执行机构,增加了机构系统方案的复杂性。因此工艺动作过程分解时的巧妙性是紧密结合实际情况,充分应用积累的知识和经验密切相关的。

#### (2) 在选择执行机构时应注意简单灵巧

要实现某一执行动作或若干执行动作选择机构时应注意利用机构的工作性能、结构特点、适用范围等主要特性,使机构造型做到合理、简单、灵巧。例如:为了分送圆柱形工件,利用摇杆和连杆的特殊形状和相对运动关系的简单铰链四杆机构来实现。如图47.5-4所示分送工件的铰链四杆机构。利用摇杆CD上的手掌形与连杆BC上的挡爪,先在位置I上接住圆柱形工件;后在位置II上传送圆柱形工件,并用挡爪挡住工件斗内其他圆柱形工件;再在位置III上将圆柱形工件落入存工件处。然后再从位置III返回位置I。如此周面

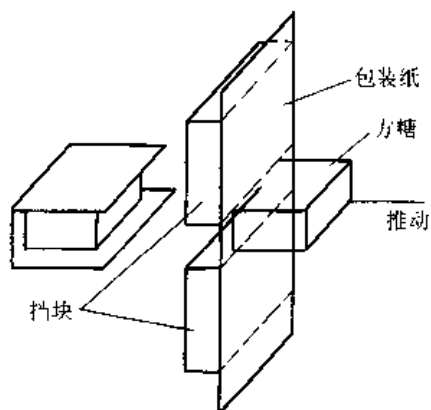


图 47.5-3 方糖折叠式包装

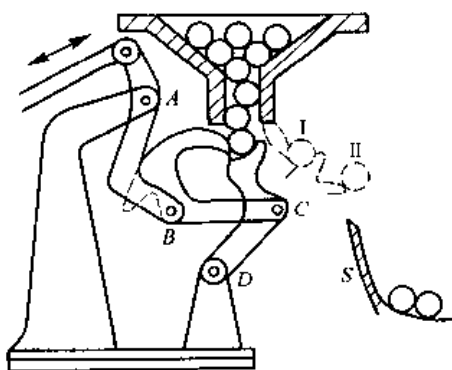


图 47.5-4 分送工件的铰链四杆机构

复始,将一个个的工件分送至存工件处。选择这种简单、灵巧的铰链四杆机构,将接工件—运送工件—存工件 3 个动作巧妙地完成,大大简化了机构系统。选择简单、灵巧的执行机构是与设计人员熟悉机构特性、富有实际经验分不开。因此,设计人员平时应多阅读相关的

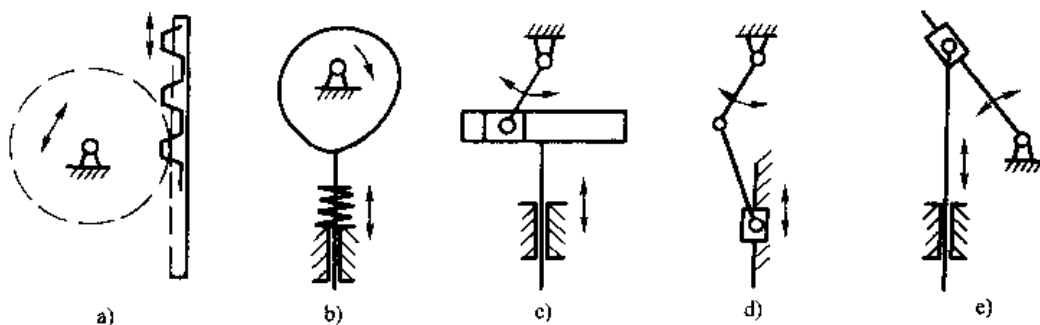


图 47.5-5 实现往复移动的机构

## 2 机械的工艺动作过程的构思

### 2.1 工艺动作过程是功能和功能原理方案的具体体现

机构设计手册。如由邹慧君主编的《机械原理课程设计手册》(高等教育出版社,1998)等等。

(3) 应采用适合于机械运动系统方案的评价指标体系,进行从机构到机构系统的评价选优

在数量较多的机械运动系统可行方案中,选择综合最优的方案并不是一件容易的事。选择合理的、可靠的、较为客观的评价指标体系和评价方法是十分重要的。对于机械运动系统来说,评价指标体系的确定应来自于有丰富设计经验的专家。否则会影响确定方案的合理性和可靠性。关于评价选优详见第 7 章。

### 1.4 机械系统运动方案的尺度综合

机械运动系统方案设计不但应包括机构系统中各机构的类型确定和各机构的运动循环图设计,而且应包括各机构的尺度综合。其实只要机器的工艺动作过程确定以后,机械运动循环图就不难确定。在机械运动循环图中往往已经决定了各执行机构的执行构件的运动类型、运动规律以及主要特性。大家知道,执行构件的运动规律不仅取决于机构的类型而且还取决于机构的尺度参数。因此我们必须对机构系统中的各机构的尺度进行设计,使其能够实现所需的运动类型、运动规律和基本特性。

例如为了满足实现往复移动的机构,可以有齿轮-齿条机构、直动从动件盘形凸轮机构、曲柄滑块机构、正弦机构、正切机构等,如图 47.5-5 所示。

但是如果使往复移动实现简谐运动规律,只有采用直动从动件盘形凸轮机构和正弦机构两种。根据简谐运动的具体要求设计出凸轮轮廓曲线和正弦机构的曲柄长度。

机器的工艺动作过程是指实现机器功能原理方案的一系列相关的动作。这些动作是按一定的时序依次完成的。

例如,牙膏装盒机是将灌装好的塑料软管装入纸盒后封口。图 47.5-6 为牙膏装盒机工艺动作初步构想

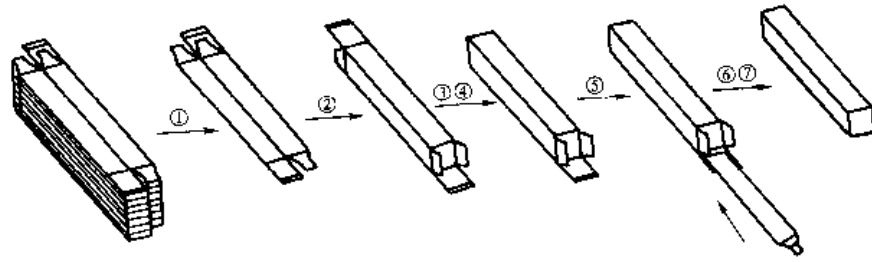


图 47.5-6 牙膏装盒机工艺动作构想

方案,它是牙膏装盒功能的功能原理方案的具体体现。它可以分为7个动作。

- 1) 将纸盒塑料坯从储存器内送出一个料坯;
- 2) 将纸盒料坯成型;
- 3) 将纸盒后面两侧舌片闭合;
- 4) 将纸盒后面盒盖封口;
- 5) 将已灌满的塑料牙膏推入纸盒;
- 6) 将纸盒前面两侧舌片闭合;
- 7) 将纸盒前面盒盖封口,装盒工作完成。

从上面牙膏装盒机工艺动作构想来看,仅仅只有一个抽象的功能——装盒,则无法进行功能求解,必须将功能分解细化,用若干有序的动作将功能具体化。

同时,上述有序动作不是一成不变的。如果采用塑壳牙膏垂直落入盒内的构想,则必须在动作之前将纸盒翻转90°,然后才有可能将塑壳牙膏垂直装入盒内。而后面原来5)将塑壳牙膏推入纸盒的动作即可取消。这样也就改变了工艺动作过程。

又如,汽车起动电机电枢绝缘纸自动嵌纸机的功能是将绝缘纸自动嵌入起动电机的电枢槽内,如图47.5-7所示起动电机电枢示意图。要完成嵌纸功能必须完成相应的若干动作,如图47.5-8所示。

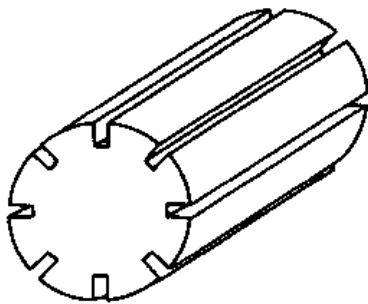


图 47.5-7 起重电机电枢示意图

它将动作分解为送纸—切纸—插纸—推纸—分度后进入下一轮嵌纸。再如,工业平缝机的功能是形成锁式线迹缝制所需缝料。它的线迹如图47.5-9所示锁式线迹示意图。图47.5-10所示为工业平缝机缝制动作的构想。其主要动作要求分为5个步骤。它们由刺料—供线—勾线—收线—送料来简单描述。

总之,功能是机器功用的描述。功能原理是实现功

能所采用的工作原理,是功能的具体实现方式。工艺动作过程则是机器工作原理实施所采用的一系列的动作。因此,构想机械的工艺动作过程是进行机械运动系统方案设计的关键内容。

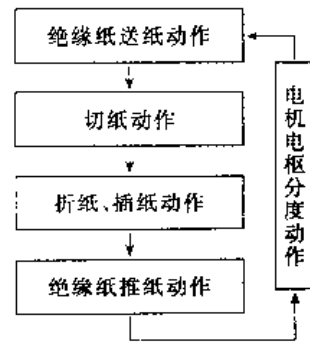


图 47.5-8 电机电枢绝缘纸自动嵌纸动作构想

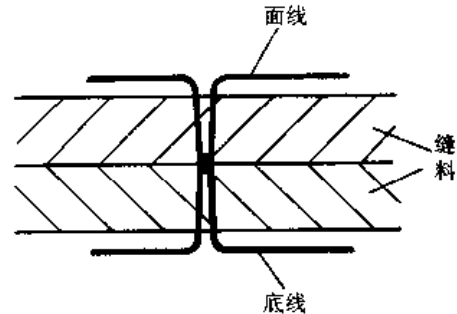


图 47.5-9 锁式线迹示意图

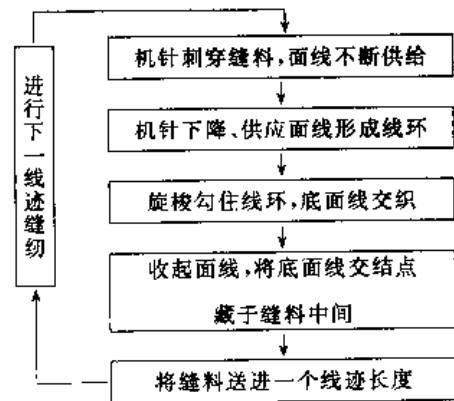


图 47.5-10 工业平缝机缝制动作构想

## 2.2 工艺动作过程与机器类型的关系

工艺动作过程与机器本身的工作机理、工作特性、工作范围、精度要求、生产率要求等等均有十分密切的关系。

对于工作机器来说又可分为下列几种类型：

### (1) 金属冷加工机械

如金属切削加工机械、冲压机械等。它们的主要功能是加工（切削或冲压），辅助功能是上料和完成加工后工件的下料。上料—夹紧—加工—松开—下料是它的主要工艺动作过程。对于不同种类的加工机械，它的加工过程是有较大差别的。加工动作可能还应细分。

### (2) 轻纺机械

如纺织机械、缝制设备、制鞋机械等，它们的主要功能是织布、缝纫，辅助功能是供线、供料和成品送出。对于缝纫机刺料—供料—勾线—收线—送料是它的主要工艺动作过程。

### (3) 包装、食品机械

如包装机械、食品机械等，它们的主要功能是包装，辅助功能是被包装物和包装物的供料和包装后物料的输出。包装物和被包装物的供料—包装—包装后成品输出是它的主要工艺动作过程。具体包装过程还应根据被包装物形态，包装物的材料等来加以细化。

### (4) 印刷机械

如印制机、装订机等，它们的主要功能是印刷、装订，辅助功能是纸张、油墨（或书贴、装订钉）的输送和成品的输出。对于印刷机它的纸张、油墨输送—印刷—成品输出是它的主要工艺动作过程。印刷的动作还取决于印刷工作原理。

### (5) 传输机械

如物料输送机械等，它们的主要功能是将工件由一个位置输送至另外一个位置，辅助功能是工件供给和输出到位。因此它的主要工艺动作过程为供给—输送—输出，相对比较简单。

除了这 5 种机械外，还有建筑机械、商业机械等等。在工作机器中，轻纺机械、食品包装机械、印刷机械等的工作特点：工作对象不是刚性的物质，工作对象往往是两种，工作原理比较复杂。因此这些类型的工艺动作过程相

对也就比较复杂。

## 2.3 工艺动作过程构想原则

工艺动作过程是机器功能具体实施所表达的动作方案，优良的工艺动作过程是设计高质量、创新性强的机器的关键所在。构想机器工艺动作过程，应遵循如下的一些原则：

### (1) 工艺动作过程必须满足机器工作原理

机器的工作原理就是实现机器功能的工作原理和方法。机器工艺动作过程就是机器工作原理的具体动作和运动过程。工艺动作过程必须满足机器的工作机理。离开机器工作机理的工艺动作过程无法得到性能优良、高效可靠的工作。

例如，构想印刷机的工艺动作过程就应懂得印刷工作机理。印刷工作机理有轮转式印刷和平版式印刷两种。因此，可以构想出这两种不同的工艺动作过程。图 47.5-11 所示为轮转式和平版式印刷机的工作原理，动作应满足机器的工作原理需要。

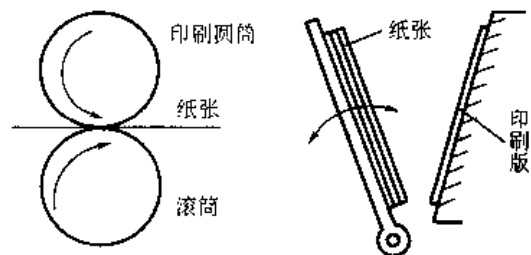


图 47.5-11 印刷机理

又如，构想缝纫机的工艺动作过程就应懂得缝纫工作原理。缝纫中所实现的线迹主要有两种：锁式线迹和链式线迹。如图 47.5-12 所示。这两种线迹的工艺动作过程是不同的。应该在工艺动作过程构想中满足线迹形成的需要。

### (2) 工艺动作过程所实现的各分功能具有相互独立性

工艺动作过程是实现机器的总功能分解后的各个分功能，一个好的机械运动系统方案，其各分功能是相互独立，使机械运动系统方案简洁、有效。

例如，印刷机的总功能分解后可得纸张、油墨输送、印刷及成品输出 3 个分功能，对应的工艺动作为纸

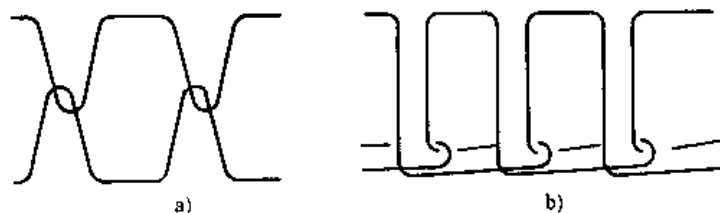


图 47.5-12 锁式线迹和链式线迹

a) 锁式线迹 b) 链式线迹



张、油墨输送动作、印刷动作和成品输出动作。如图 47.5-11 所示,印刷机的分功能和工艺动作过程。分功能的相互独立使工艺动作互不干涉,并使工艺动作过程构思简单明了。

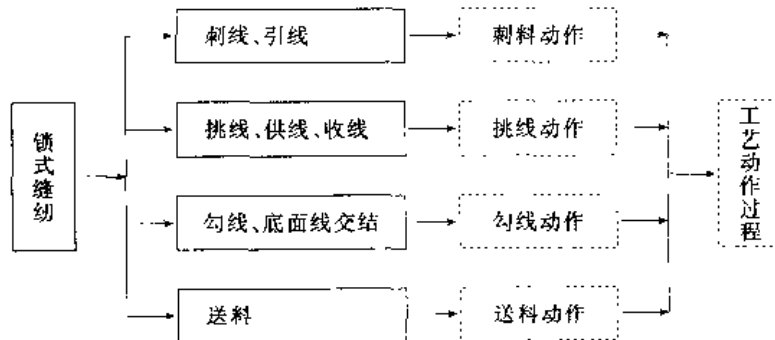


图 47.5-13 锁式缝纫机的功能分解和动作过程

(3) 工艺动作过程的各动作要易于为常见的执行机构实现

工艺动作过程最终是由各执行机构的执行构件运动来实现的。机器的工作原理,功能的分解,工艺动作过程的构想均应考虑最后能易于为常见的执行机构来实现。常见的执行机构能实现等速和不等速转动、往复摆动、往复移动,间歇移动、间歇摆动、刚体导引、实现函数、实现轨迹以及复合运动等等。通常的机械运动系统均由这些常见的执行机构组合而成。如果工艺动作过程中的某一动作不能由常见的执行机构来完成,那么应考虑采用实现运动比较复杂的组合机构,或者采用两个执行机构来完成这一动作。

有时,某种工艺动作过程的某些动作找不到合适的执行机构来实现,此时就无法设计出实现此工艺动作过程的机械运动系统方案来。例如,在 19 世纪中叶,人们为了提高缝制衣服的速度,致力于缝纫机的发明工作。开始时,人们考虑用机器来完成拟人的缝纫动作过程,也就是用一根针、一条线、线穿在针的尾部来进行缝纫。由于当时没有找到实现这种缝纫动作的相应的执行机构,致使发明缝纫机的工作宣告失败。但是后来有人采用了锁式线迹的缝纫动作过程,见图 47.5-12a 所示,这种缝纫动作过程是采用一根针、两条线、面线穿在针头孔中,底线绕在梭心中,通过刺布机构、挑线机构和勾线机构进行底面线交织来完成缝纫。由于机构的发展,人们最终还是发明了撬缝缝纫机,实现了拟人缝纫。

20 世纪 70 年代开始,机电一体化技术的不断完善,使传统的机构学发展成现代机构学。现代机构的主要特点是采用单自由度以及多自由度的可控机构,使机构能实现复杂多变的动作,因此使机械运动系统设计有了更广阔的途径。

又如,锁式缝纫机的总功能分解后分功能如图 17.5-13 所示。可见分功能是相互独立的,可用对应的工艺动作加以实现,这些工艺动作按一定的时序组合成锁式缝纫机的工艺动作过程。

## 2.4 工艺动作过程的构思方法

机械运动系统设计的一个关键问题是构想出一种简单、易行、高效的工艺动作过程。在构思工艺动作过程前必须懂得机器的工作机理。不懂工作机理便无法构想出好的工艺动作过程。在构思工艺动作过程前还必须熟悉执行机构的类型、运动特点和工作性能等,否则会使构想工艺动作过程陷入困境。

常用的构思工艺动作过程的方法有 3 种:

### (1) 拟人动作过程的构思方法

不少工作机器均是为了代替人类的手工劳动。例如:纺织机械、包装机械、食品加工机械等等。它们均可按人工加工过程中的动作来构思机械的工艺动作过程。

为了构思医用棉签卷棉机的工艺动作过程,可以先去观察护士医用棉签卷棉的过程:先将医用棉分段取出,再将牙签取出,然后将医用棉包在牙签上,最后将牙签转动而形成棉签。只要实现上述过程就可构思出棉签卷棉机的运动系统方案来。

为构思书本包装机的工艺动作过程,可以先观察人工包装的动作过程:先取出包装纸,后将书本放在纸上,再将两侧包装纸折上,最后将两端包装纸折合好。这种动作过程就是书本包装机的工艺动作过程。

拟人动作的方法要注意两点:一是要将多种可能的拟人动作进行精心的挑选,以求高效;二是选择好易于为机构实现的拟人动作,以求易于机械化。

### (2) 物料流引发动作过程的构思方法

任何工作机器主要功用是进行物料的运动变化和形态变化。通常工作机器物料流所引发的动作过程为:相关物料输送到位—相关物料形态变化—新形态物料的输出。

这种物料流所引发的动作过程连贯在一起就可成

为较为自然合理的工艺动作过程。

例如扭结式糖果包装机的物料流动作过程，见图 47.5-14 所示，它由包装纸、糖果输送到位，糖纸裹包扭结包装和包装后糖果输出等 3 个物料流动作组成。

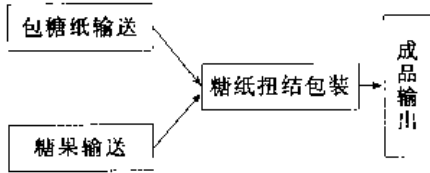


图 47.5-14 扭结式糖果包装机物料流

再如，核桃剥壳机的物料流动作过程，见图 47.5-15 所示，它是由核桃到位、核桃挤压破壳以及核桃壳和核桃肉输出等 3 个物料流动作组成。

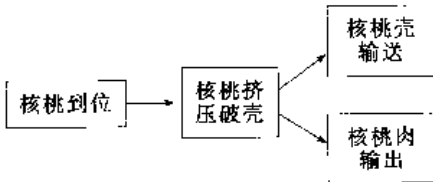


图 47.5-15 核桃剥壳机物料流

(3) 基于功能细化和分解的工艺动作过程构思方法

功能细化和分解将会有利于对工艺动作过程的构思。第六章我们提出了通过抽象化可以得到机械的“核心功能”。但是作为一台完整的机械功能，还应包括“前导功能”和“后续功能”。如图 47.5-16 表示基于功能细化和分解的工艺动作过程的构思。

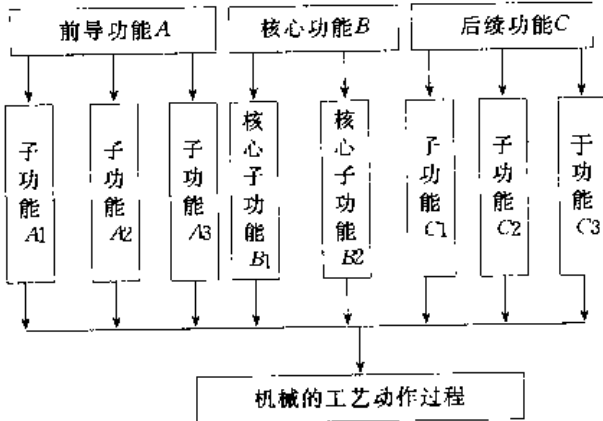


图 47.5-16 基于功能细化和分解的工艺动作过程构思

前导功能是将所需要物料输送到位。核心功能是机械所应完成的、物料的运动和形态的变化。后续功能是将完成关键功能的物料输出。将完成这 3 个功能所分解出的所有子功能的系列行为动作连贯在一起，可以构成机械工艺动作过程。

例如，平版印刷机的核心功能是在白纸上印出文字和图案。它的前导功能是上油墨、在铅字版上刷油墨和将白纸输送到位。它的后续功能是将铅字版和印刷

完的纸张分离并取出印刷完毕的纸品。它的工艺动作过程构思如图 47.5-17。

基于功能细化和分解的方法，实质上是功能分解法，将各子功能寻求它的动作解，再来构思出工艺动作过程。

从上述几种构思工艺动作过程来看，这是机械创新设计的重要步骤。同一机械所要实现的工艺动作过程可以是不同的，不同的工艺动作过程就会引发出一种全新的机械运动系统方案。

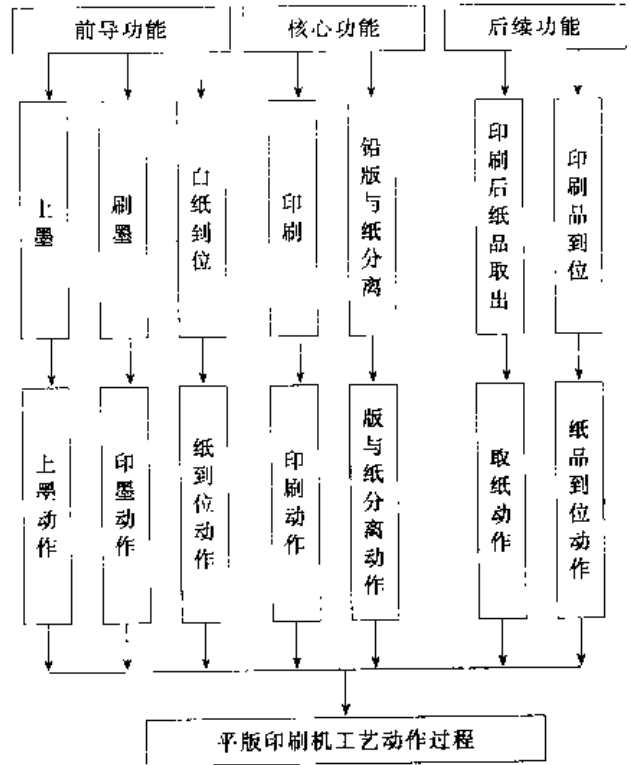


图 47.5-17 平版印刷机工艺动作过程构思

### 3 机械工艺动作过程分解和执行机构的选择

#### 3.1 机械工艺动作过程的内涵

机械工艺动作过程是反映机械的工作原理和工艺特点，直接影响了机械的创造性和工作特性。具体来说机械工艺动作过程应包含下列内容：

(1) 从坯件到成品的具体工作过程

对于工作机器来说，坯料就是提供的毛坯或原始物料。成品是指毛坯经过形态变换后的产品，或将原始物料经过组合后的产品，表达了具体的工作过程。

(2) 表示产品工艺动作顺序和工位数

工艺动作过程表示完成总功能的各个工艺动作的先后顺序，反映了动作协调配合的工作过程。从机械运动系统方案设计出发，可以将机械的工艺动作过程用

若干工位来完成。每个工位所完成的工艺动作情况决定了应该有多少工位。工位较多,每个工位完成动作可以较为简单。但是会使机械运动系统方案由于工位增多而复杂化。相反,工位较少,每个工位完成动作会趋向复杂,使机械运动系统方案由于工位减少而得到简化。因此,合理地选择工位对于机械运动系统方案设计是至关重要的步骤。

(3) 表示产品在各工位上的加工状态及运动形式

一旦由工艺动作过程分解成若干工位,每个工位上的加工状态和运动形式即取决于此工位所需完成的功能。加工状态取决于机械工作机理的需要,运动形式取决于物料和工作需要。

(1) 表示产生动作的构件所处位置、运动方向和运动形式

产生动作的构件又称执行构件。执行构件的运动方向和运动形式对于设计机械运动系统方案是十分重要的。因为执行构件的运动方向、运动类型、运动规律是我们选择执行机构的依据。

### 3.2 机械工艺动作过程的分解

机械工艺动作过程的分解是机械运动系统方案设计中十分重要的步骤。通过合理的分解才能得到各工

位上的运动形式和运动规律,才能确定相应的执行机构。

(1) 机械工艺动作过程的分解准则

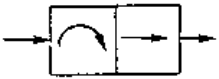
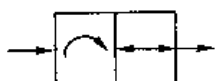
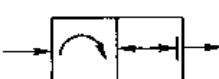
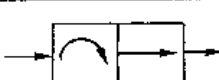
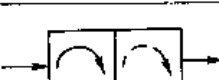
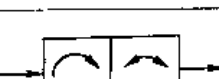
为了使机械工艺动作过程分解更趋合理,应满足下列分解准则:

1) 动作最简化原则 在机械工艺动作过程分解时应遵循最简化原则,即采用最简单的一系列动作组成此工艺动作过程。分解动作愈简单,将来采用对应的执行机构也愈简单。

2) 动作可实现性原则 在机械工艺动作过程分解时应遵循可实现性原则,由于在机械中任何动作都由执行机构来实现,而常用的执行机构可以实现的动作是有限的,如表 47.5-1 所示常见的运动形式变换中输出运动有 9 种。分解后的动作基本上应属于上述 9 种之内,这就符合可实现性原则,否则要采用组合机构或可控机构来实现,增加了设计工作的难度。

3) 动作数最小原则 在机械工艺动作过程分解时应遵循分解后所得动作数尽量减少的原则,动作数的减少可以使执行机构数目减少,从而使机械运动系统方案简化。图 47.5-18 所示,只要上、下加固定的板块,用一个移动的动作推动被包装物,就可实现三面的包装,有利于机械运动系统的简化。

表 47.5-1 常见执行机构输出动作的形式、符号、实现机构

序号	运动形式变换内容	符 号	实现功能的机构
1	连续转动变单向直线移动		内齿轮条机构、螺旋机构、蜗杆齿条机构、带传动机构、链传动机构等
2	连续转动变往复直线移动		曲柄滑块机构、移动推杆凸轮机构、正弦机构、正切机构、牛头刨机构、不完全齿轮齿条机构、凸轮连杆组合机构等
3	连续转动变带停歇往复直线移动		移动推杆凸轮机构、利用连杆轨迹实现带停歇运动机构、组合机构等
4	连续转动变单向间歇直线移动		不完全齿轮齿条机构、曲柄摇杆机构+棘条机构、槽轮机构+齿轮齿条机构、其他组合机构等
5	连续转动变单向间歇转动		槽轮机构、不完全齿轮齿条机构、圆柱凸轮式间歇机构、蜗杆凸轮间歇机构、平面凸轮间歇机构、内啮合星轮间歇机构等
6	连续转动变双向摆动		曲柄摇杆机构、摆动导杆机构、曲柄摇块机构、摆动推杆凸轮机构、电风扇摇头机构、组合机构等

(续)

序号	运动形式变换内容	符号	实现功能的机构
7	连续转动变带停顿双向摆动		摆动推杆凸轮机构、利用连杆曲线实现带停顿运动机构、曲线导槽的导杆机构、组合机构等
8	往复摆动变单向间歇转动		棘轮机构、钢球式单向机构等
9	连续转动转变为实现预定轨迹		平面连杆机构、连杆-凸轮组合机构、联动凸轮机构、精确直线机构、椭圆仪机构等

实现动作数量少，要求我们针对工艺动作过程认真分析、精心构思。

(2) 机械工艺动作过程的分解方法

机械工艺动作过程的分解方法可以归纳为以下 3 种：

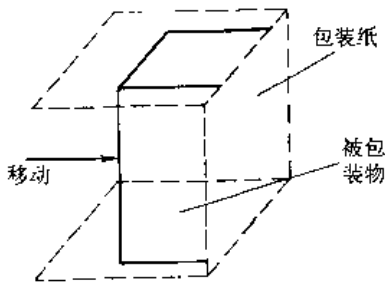


图 47.5-18 同时包装三面的一个动作

1) 物流运动状态表示法 机械工艺动作过程是物流运动状态的具体描述。利用物流状态的具体化过程，可以得出相应的运动动作，分解出若干个执行动作。例如：图 47.5-13 的锁式缝纫机的工艺动作过程，它的物料运动状态必须完成：刺料引线—供线收线—勾线—送料 4 个执行动作。又如，图 47.5-17 的平版印刷机的工艺动作过程按它的物料运动状态必须完成：纸墨输送—印刷动作—印刷品输出等执行动作。

由此可见物流运动状态的表达是机械功能的具体实现过程。

2) 功能-行为法 (F-B) 法 机械工艺动作过程是为了实现机械的总功能。因此，采用功能-行为法可以对体现机械工艺动作过程的总功能进行分解，将分解所得的分功能求得相应的动作行为，从而实现工艺动作过程的分解。如图 47.5-19，表示了功能-行为法的

具体分解过程。由此可见不同的总功能分解方法，可以求得不同的机械工艺动作过程，也就会得到不同的动作行为。

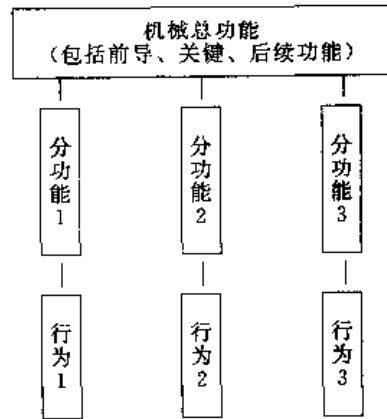


图 47.5-19 功能-行为法求解动作

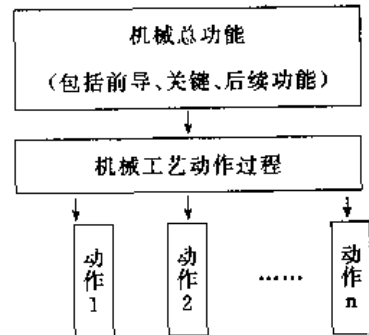


图 47.5-20 功能-动作过程-动作法

(3) 功能-动作过程-动作法 (F-P-A 法) 由机械的总功能构思机械的工艺动作过程，然后将工艺动作过程进行动作分解得到相应的各个执行动作，其过程

可以用图 47.5-20 来表示。例如,扭结式糖果包装机,它是要求将糖纸裹包糖果后将糖纸扭结包装。其物料流见图 47.5-14 所示。上述扭结包装糖果的工艺动作过程可以分解成下列动作:输送糖纸—输送糖果—糖纸裹包糖果—糖纸扭结—糖果成品输出。

工艺动作过程的分解对于机械运动系统方案构思和设计是十分重要的。如何分解好机械工艺动作过程是需要借助于对此类机械工作机理的认识和设计者的设计经验。同一机械工艺动作过程可以有不同的分解结果,因而可以构思和设计出不同的机械运动系统方案来。

上述机械工艺动作过程的分解准则和分解方法将会有利于我们做好机械工艺动作过程的分解。

### 3.3 动作组合的创新

从机械运动系统方案设计的流程来看,由动作来映射执行机构是机械运动系统方案设计中具有创新作用的重要环节,也是目前实现计算机辅助创新的关键步骤。

计算机辅助机械运动系统方案设计的主要任务是按实现动作的需要寻求大量的、符合基本条件的可行的执行机构系统方案。在方案评价标准一定的条件下,计算机产生的可行方案越多,其最终得到的最优方案的质量通常也越高。

为了使产生的可行方案数量增加,我们提出了针对动作-机构(A-M)的创新方法:

#### (1) 动作分组法

动作直接映射为机构,还需要解决两个问题:

1) 如何确定执行机构的输出运动。一个较为复杂的动作可以分解为若干个简单动作;反之,某些简单动作也可以合成为一个较复杂动作。但是哪种输出运动更有利于实现机械的功能和关键。

2) 如何确定机械运动系统应有多少执行机构,应该考虑多少执行机构组成整个系统才是最佳的。

为了解决上述两个问题,我们提出两个措施:

1) 规定一组动作能且仅能描述一个执行构件的运动;

2) 假定一个执行机构只有一个执行构件。

动作分组法基本思路,如图 47.5-21 所示。表示将复杂动作分解为较为简单的动作,也可将相互矛盾、不可共存的两个动作  $A_1$ 、 $A_2$  加以分开。若动作小组 1、2 中动作仍不可共存就继续分组,直至动作小组数等于简单动作个数。通过这种办法,可遍历所有可能的动作组合情况,完成对机械运动系统动作过程的组合解的全面搜索,有利于进行机构系统的创新。

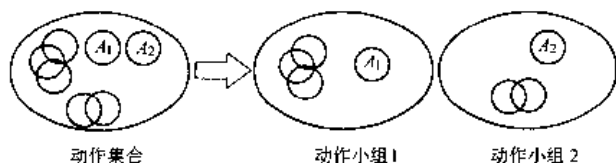


图 47.5-21 动作分组法

为了判别动作是否可共存,其判断规则(又称动作集合分组规则)如下:

- 1) 串连的简单动作可以共存;
- 2) 动作并联,若其运动形式相同,方向相反,则不能共存;
- 3) 动作并联,若动作之间存在相对运动关系,则不能共存;
- 4) 若物料不可穿越,运动范围在物料两侧的动作,不宜共存;
- 5) 运动轴线相互垂直的动作,不宜共存;
- 6) 作用力相差悬殊的动作,不宜共存;
- 7) 其他导致动作不宜共存的情况。

除了上述 7 条动作集合分组规则外,还需判断分组是否合适的分组评价规则:

- 1) 若估出的小组动作的运动总时间远大于执行机构系统的工作循环时间,则认定分组不合适;
- 2) 若无法找到实现小组动作的执行机构,则认定分组不合适;
- 3) 若不能满足功能的其他要求(除运动规律外),则认定分组不合适。

动作分组法的具体步骤见图 47.5-22 动作分组法流程图。

在图 47.5-21 所示的动作集合中共有  $n=7$  个行为,其中两个动作不可共存。若该集合仅分为两组,则可行动作分组数可按下式计算:

$$C_7^0 + C_7^1 + C_7^2 + C_7^3 + C_7^4 + C_7^5 = 32$$

由此说明这种动作集合可以由 32 个可行动作组合方案。用动作分组法可以实现机构系统方案的创新。

#### (2) 动作变换法

动作变换法是通过改变动作的表现形式来创新新的动作组合形式。动作变换主要包括动作合并、动作分割、动作分解和动作分位等形式。

1) 动作分割和合并 动作分割是将一个动作分割成若干子动作;动作合并是将若干动作合并成一个动作。

图 47.5-23 表示将一个动作分割成若干个串连的子动作。图 47.5-24 表示将若干个并联的子动作合并成一个动作。不同的动作分割方式,对生产率的影响是不同的。

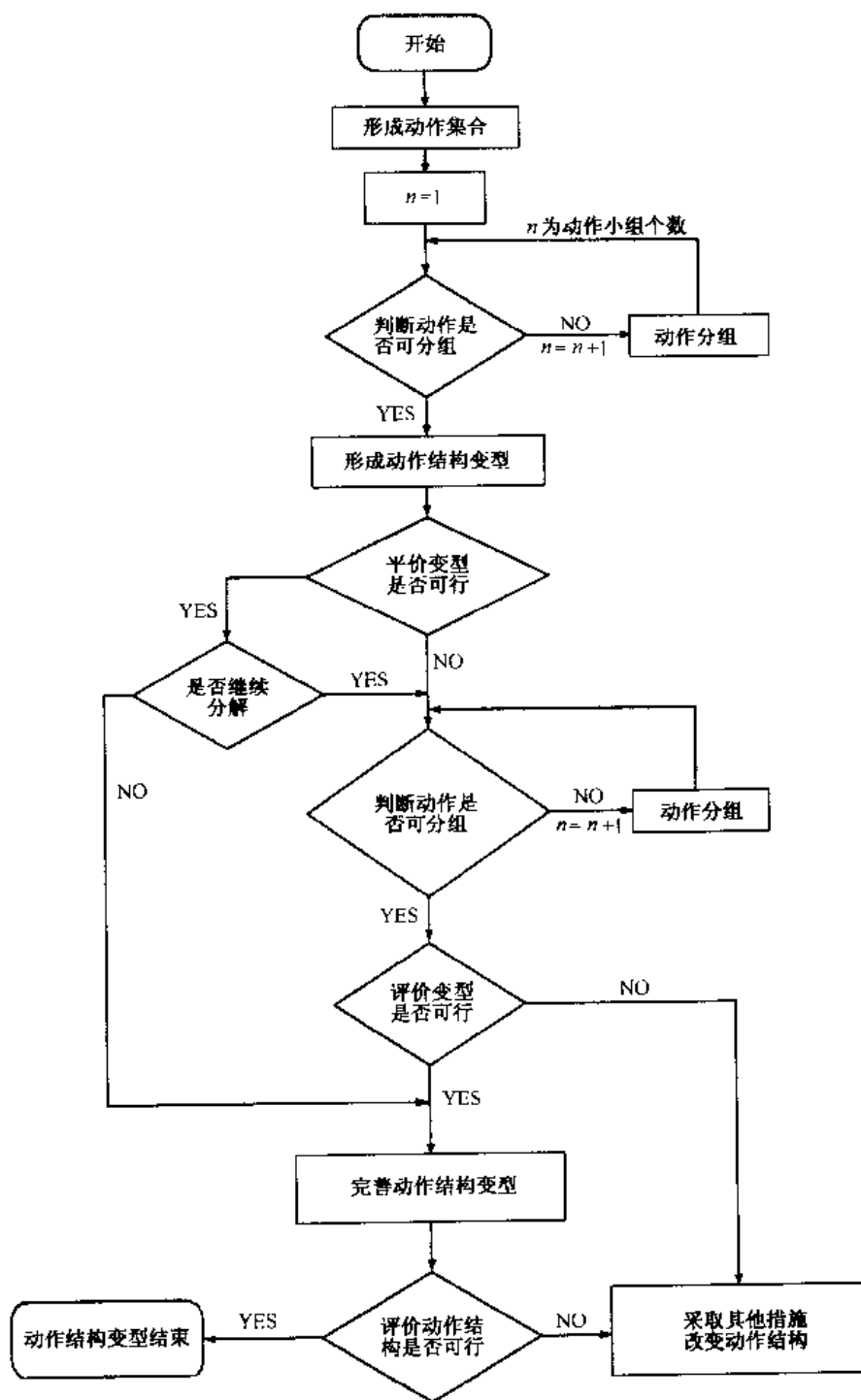


图 47.5-22 动作分组法流程图

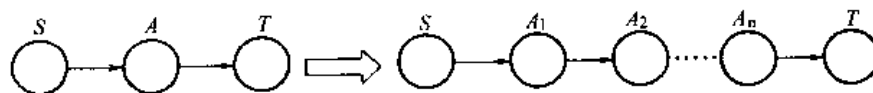


图 47.5-23 动作分割

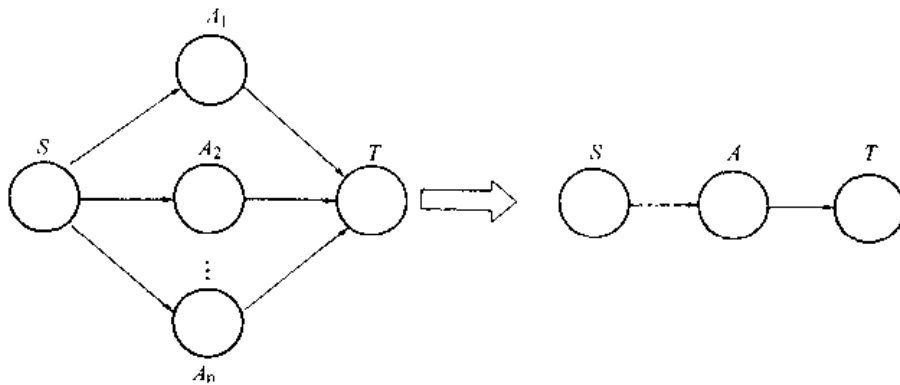


图 47.5-24 动作合并

2) 动作的分解 动作分解是按一定的分解原理, 将一个动作分解为两个或两个以上的子动作。

图 47.5-25 表示实现印刷功能的动作分解的若干方案, 动作分解可产生许多可行的方案。因此, 动作分解就会有创新方案的可能性。

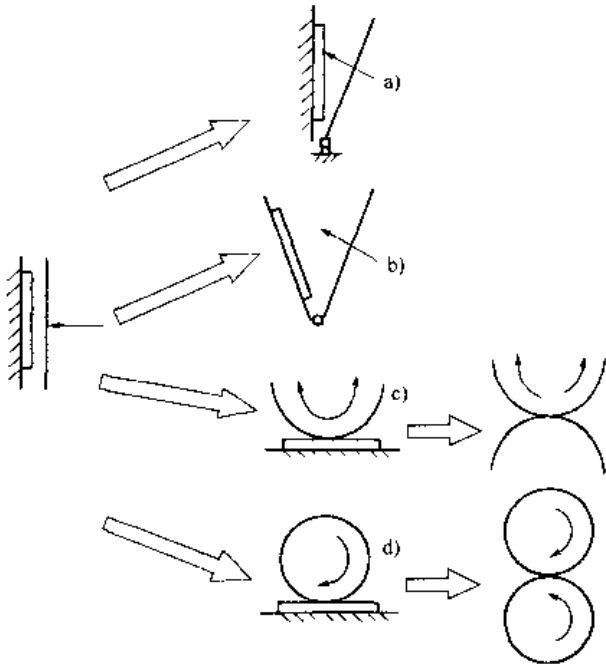


图 47.5-25 实现印刷功能的动作分解

3) 动作的分位 动作分位就是将一个动作分若干个工位来完成。动作分位的目的是提高生产率, 方法是分配动作到适当的空间位置上。对于动作是否可以分位及如何分位, 我们先做一个假设:

在一个工作循环中, 物料经过一个工位只有一次。动作是否可分位的判别规则如下:

1) 存在相对运动关系的两组动作, 无法分在不同工位上;

2) 两组动作之间存在一个时序关系, 且发生时序关系的两动作不是各自组的最后一个动作, 也不是各

自组的最前一个动作, 则无法分在不同工位上;

3) 两组动作之间存在不止一个时序关系, 则两组动作无法分在不同工位上;

4) 两组动作之间存在一个时序关系且发生时序关系的两动作中, 先完成的动作是该组的最后一个动作或后完成的动作是该组的最前一个动作, 则这两组动作可分在不同工位上。

### 3.4 动作的描述和机构属性表达方式分析

在机械运动系统方案中, 采用执行机构来实现各个执行动作, 执行机构是通过运动形式、运动方向和运动速度的变换, 运动的合成和分解, 运动的缩小和放大以及实现给定的运动位姿和轨迹等来表达功能元的功能。机械运动系统的工艺动作过程往往是由一系列复杂运动来完成的, 这些复杂运动又可看成是由一系列简单的动作 (或称基本运动), 如单向转动、单向移动、往复摆动、往复移动、间歇运动等的组合而成。因此, 利用运动转换功能图, 便于找出与要求的运动特性相匹配的机构, 使机构造型过程更具直观性。但是在机构选型和组合过程中, 还应考虑运动轴线、运动速率的变化。

图 47.5-26 表示了四工位专用机床中从动力源到机床各执行件 (安装工件的工作台、装有刀具的主轴箱和刀具) 之间的运动类型的转换和速率大小的变化。但是, 没有反映出从动力源到机床各执行件的轴线位置变化情况, 两个电动机的回转轴线是一个未知量, 3 个执行件 (工作台、主轴箱和刀具) 的运动轴线也不明确, 而在机械运动系统方案设计中, 动力源和各执行件的轴线位置恰恰是一个十分重要的因素, 有时甚至是决定方案是否可行的一项指标。此外, 图中各运动速率特性 (匀速、匀加速还是非匀速等等) 也没有明确表示出。

在机械运动系统方案设计中, 设计者最关心的机构运动功能是输入和输出的运动形式 (也称动作类型)、运动轴线、运动方向及运动速率等。

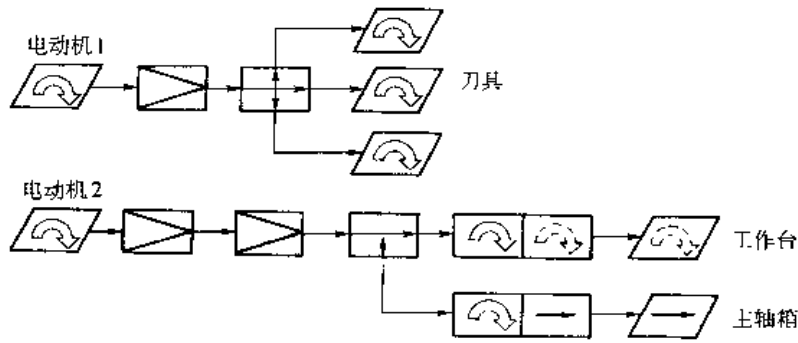


图 47.5-26 四工位专用机床  
运动转换功能图

机构的运动特性可以概括为：运动形式的转换、输入和输出轴的方向、输入和输出的速率大小。

图 47.5-27 表示机构运动特性的描述。

运动形式的转换具体表现为：输入和输出运动形式间的变化。对于机构来说，输入运动主要有：转动、摆动或移动；输出运动主要有：单向转动、单向移动、

往复摆动、往复移动、间歇摆动、间歇移动、刚体导引运动、点的轨迹运动等等。

输入和输出的基本运动方向可具体描述为：绕 3 个坐标轴的单向或双向转动与摆动；沿 3 个坐标轴的单向或双向移动。

输入和输出的速率大小包括速率特性。

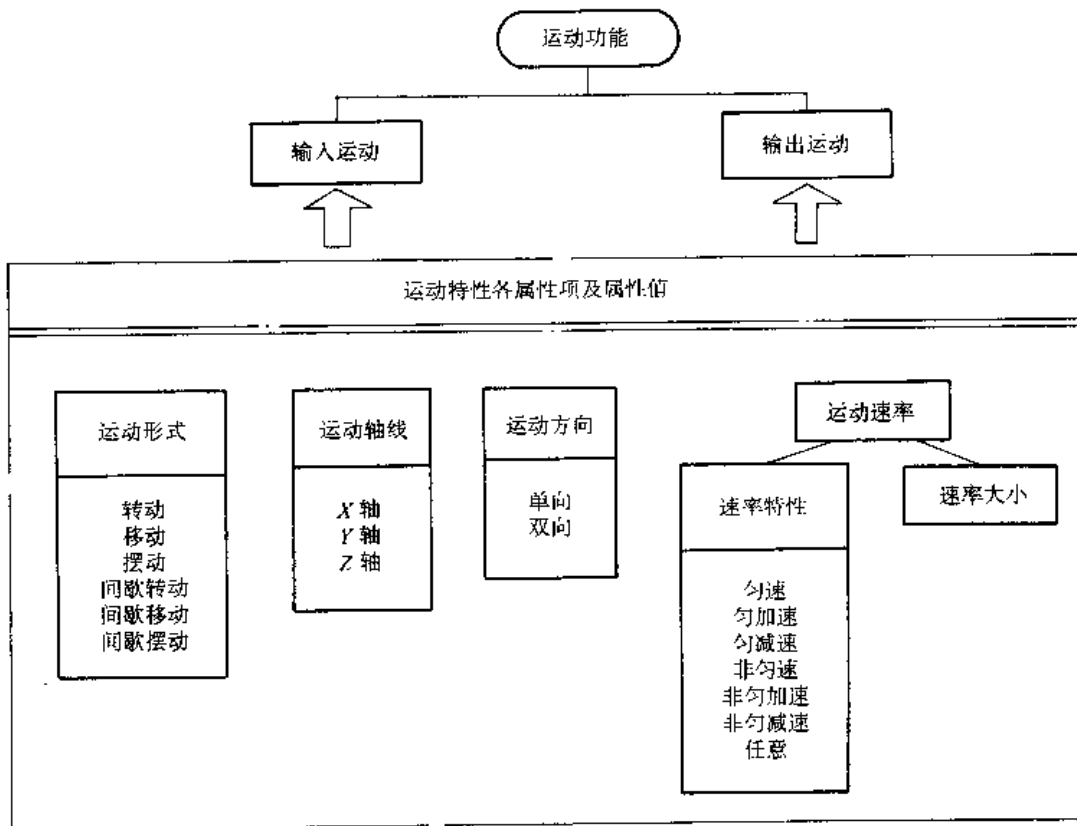


图 47.5-27 机构运动特性的描述

### 3.5 执行机构的选择

对于单自由度机构，其输入和输出构件都有确定的运动形式，表 47.5-2 表示常见机构输入、输出运动形式。

为了选择好合适的执行机构，还可参见贝季瑶主编的《现代机械设备设计手册》中第 8 篇“机构及其系统设计”（机械工业出版社，1996），邹慧君主编的《机械原理课程设计手册》（高等教育出版社，1998）。



表 47.5-2 常见机构输入、输出运动形式

序号	常见机构	输入运动形式	输出运动形式	传递运动特性
1	圆柱齿轮机构	等速转动	等速转动或移动	传递两平行轴转动
2	螺旋齿轮机构	等速转动	等速转动	传递两交错轴转动
3	蜗杆蜗轮机构	等速转动	等速转动	传递两垂直交错轴转动
4	锥齿轮机构	等速转动	等速转动	传递两相交轴转动
5	盘形凸轮机构	等速转动或移动	往复移动或摆动	传递平面内运动
6	圆柱凸轮机构	等速转动	往复移动或摆动	传递空间运动
7	平面连杆机构	等速转动	往复移动或摆动 刚体导引或点的轨迹	传递平面内运动
8	空间连杆机构	等速转动	往复移动或摆动	传递空间运动
9	间歇运动机构	等速转动	往复间歇摆动或移动	传递两平行轴或相交轴或交错轴运动
10	组合机构	等速转动	往复摆动或移动 刚体导引或点的轨迹	一般为平行轴间运动

#### 4 机械运动系统方案的组成原理与方法

机械运动系统方案的组成是将所选的执行机构构成若干可行的机械运动系统方案（亦可称之为机构系统的组成），它应包括机构系统的型综合和机构系统的尺度综合。型综合是确定机构的类型；尺度综合是确定所选机构的各运动尺寸。对于机构系统的型综合包括各执行机构类型的选择和它们的相互作用关系的确定。机构系统的尺度综合是根据各执行机构的执行动作的要求进行各机构的运动尺度的计算和各机构动作时间序列确定，从而使各机构的输出运动完全满足机械的工艺动作过程要求。

##### 4.1 机械运动系统组成的相容性原则

在机械运动系统中各执行机构的组成大多采用串连形式和并联形式。在组成机械运动系统方案时，它的相容性主要反映在保持各执行机构运动的同步性、各执行机构输出动作的协调性以及各执行机构输出运动精度的匹配性。

###### (1) 各执行机构运动的同步性

同步性是反映了机械运动系统在一个机械工作循环中各执行机构有相同的工作周期，按一定的节拍完成机械工艺动作要求。通常情况下，要求各执行机构的输入构件按同一转速转动或按一定的平均速比转动，使各执行机构的运动周期相同。当然，在某些特殊的机械工艺动作过程中，个别执行机构的运动周期是其他执行机构的运动周期的整数倍。那时，可以通过定传动比的机构传动来加以保证。

例如，图 47.5-28 为家用缝纫机机构系统方案图，其中刺布机构和挑线机构的输入构件均为缝纫机的上

轴，它由电机带动旋转，使刺布机构和挑线机构的运动具有同步性。摆动的勾线机构和移动的送布机构的输入构件均由上轴通过一个曲柄摇杆机构和一个凸轮式高副机构传至下面三根轴产生摆动。因此，送布机构和勾线机构也一定与刺布机构和挑线机构同步运动。

又如，图 47.5-29 所示内燃机机构系统方案图中，1'-2-3-7 为主机构——曲柄滑块机构；4'-5-7 及 4''-6-7 为吸气及排气凸轮机构；1-4 为齿轮机构。吸气及排气凸轮机构的两凸轮在同一轴上，而主机构曲柄与凸轮通过定传动比齿轮机构建立确定的运动关系，因此两者也是同步运动。

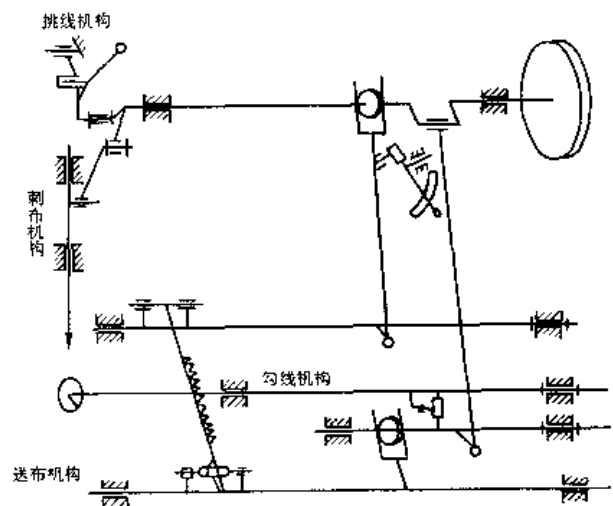


图 47.5-28 家用缝纫机机构系统方案图

###### (2) 机构输出运动的协调性

在机械运动系统方案组成时，应考虑各机构的输出运动特性。机构的输出运动特性包括：运动形式、运动轴线、运动方向和运动速率、机械运动系统设计时先

考虑采用何种工作原理；工作原理确定后需构思工艺动作过程。为了满足工艺动作需要，在组成机械运动系统方案时，应使机构的运动特性符合规定要求，使各执行机构的运动相互协调。

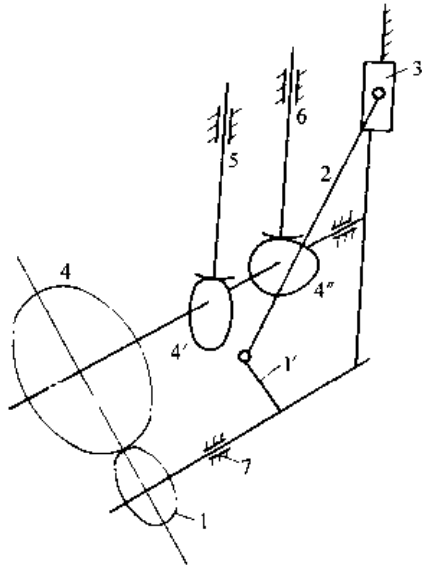


图 47.5-29 内燃机机构系统方案图

例如图 47.5-28 家用缝纫机机构系统中各机构的运动形式、运动方向和运动规律应满足刺布—挑线—勾线—送布的要求，同时运动轴线布置也是符合各机构的要求。又如，图 47.5-29 内燃机机构系统方案图中，主机机构的轴线和凸轮机构的轴线相平行，因此采用圆柱齿轮机构传动。如果主机机构的轴线与凸轮机构的轴线相互垂直，那么应采用锥齿轮机构传动，等等。

### (3) 机构输出运动精度的匹配性

在进行机械运动系统方案组成时，还应考虑机构输出运动精度的匹配。因为机械的工艺动作过程是一个整体，对于组成的各个动作精度都是有要求的。运动精度的匹配性是指各机构的运动精度能满足完成工艺动作的需要。选择过低的机构运动精度会使机械无法工作。选择过高的机构运动精度会使制造和设计成本大大提高，同时也没有必要。

## 4.2 机械运动系统组成的系统最优化原则

机械运动系统从本质上看也是一个机械系统，它的各组成部分——各执行机构是子系统，它们均需服从系统组成的基本原则和基本特性。对于机械运动系统来说，它的组成必须符合系统综合最优的原则。

为了达到系统综合最优，首先我们必须对其子系统——执行机构进行综合最优的评价和选优。对于每个执行机构必须从实现功能、工作性能、动力性能、经济性和结构紧凑等五个方面来进行评价。并由评价结

果从众多的可行执行机构中择优选择。其次我们应对整个机械运动系统—机构系统从整体的实现功能、工作性能、动力性能、经济性和结构紧凑来进行全面评价，从众多的、可行的机械运动系统中选择几个综合最优方案，供最后决策和选择。

机械运动系统追求整体综合最优方案，是系统设计的关键所在。

## 4.3 寻求执行机构的创新设计是机械运动系统创新设计的基础

要得到具有创新性的机械运动系统方案，其基础是创新出整体性能最佳的新颖执行机构。

机构的创新设计方法，已在第 3 章中加以描述。在执行机构层次上的创新设计，可以使机械运动系统方案更具创新性。

对于较为复杂的动作，如果采用常用的基本机构无法实现，可以采用组合机构和机构组合。组合机构常见的有齿轮-连杆机构、凸轮-连杆机构和凸轮-齿轮机构，它们均是将一些凸轮、齿轮、连杆的元素融合在一起的复合机构。因此它们所能实现的运动规律和运动轨迹比基本机构更加复杂多变。机构组合也是实现比基本机构更为复杂的运动规律，一般它们是采用两种基本机构的串连和并联来实现的。

## 5 机械系统运动方案设计举例

为了说明机械运动系统方案设计的步骤和大致过程，现列举两个实例。通过实例可以说明机械运动系统的构思和设计的基本原理和基本方法，同时希望能使读者起到举一反三的效果，开阔构思思路，启迪创新思维。

### 5.1 设计平版印刷机的运动系统方案

#### (1) 机器的功能和设计要求

机器的功能是表达机器的功用。机器的设计要求是机器设计的出发点。简易平版印刷机用于中、小型印刷厂，它可以印刷各种表格、联单、账簿、商标、名片等 8 开以下的印刷品。简易平版印刷机由于具有结构简单、成本低廉、使用方便、维修容易等特点，因而目前仍广泛地得到应用。

平版印刷机的功能是在小型铜锌版上刷上油墨，通过铜锌平版与白纸的相互贴合而完成印刷工艺。

为了实现平版印刷工艺，平版印刷机必须完成纸张输送、油辊上添加油墨、铜锌版上刷墨、白纸与刷墨后的铜锌版贴合印刷、取出和叠好印刷品 5 个分功能。

平版印刷机的设计要求和参数有：

- 1) 印刷能力——24 次/min；

2) 驱动电机——采用 Y90S-6:  $P=0.75\text{kW}$ ,  $n=910\text{r/min}$ ; 或 Y90L-6:  $P=1.1\text{kW}$ ,  $n=910\text{r/min}$ ;

3) 电动机安装在印刷机底部或墙板的侧面;

4) 机械运动方案应力求简单, 其固定铰链点布置在规定的墙板上;

5) 印头的固定支撑位置见图 47.5-30, 印头摆角  $50^\circ$ ; 油辊在刷墨过程中需占据的两个极限位置  $F_1$ 、 $F_2$ , 也表示在图 47.5-30 上。

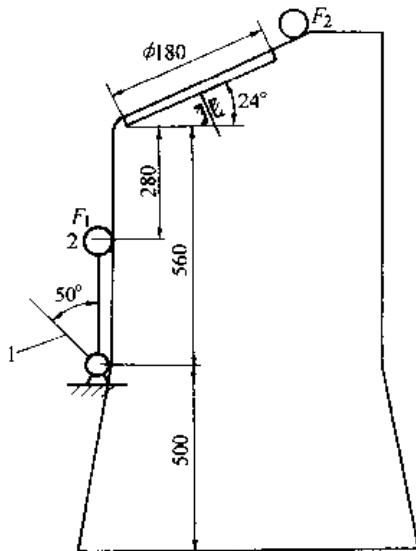


图 47.5-30 印头固定支撑位置

6) 从提高印刷质量来考虑, 希望印头在印刷的瞬间有一短暂的停歇;

7) 为了使油辊刷墨均匀, 希望油辊在工作行程和回程中的速度尽量均匀。

#### (2) 工作原理与工艺动作分解

为了实现平版印刷的功能, 可以有两种工作方式: 一是铜锌版固定, 纸张由印头带动与之贴合以完成印刷工艺; 二是纸张固定, 铜锌版由印头带动与之贴合以完成印刷工艺。这两种不同的工作方式, 它们的工艺动作过程是有区别的。下面以前一种工作方式为例说明工艺动作分解情况。对于简易式平版印刷机, 它的输纸和取出印刷品均由手工完成。因此它的工艺动作可分解为:

1) 由间歇动作机构给油辊上墨;

2) 由油辊上下运动完成在铜锌版上的均匀刷墨;

3) 由纸张来回摆动与涂墨后的铜锌版贴合完成印刷工艺。

为了使油辊上墨均匀, 先将油墨定量输送至油盘, 再将油盘定期间歇转动达到不断均匀上墨。

为了使油辊均匀刷墨于铜锌版上, 要求在刷墨时油辊尽可能等速运动。

为了使纸张与铜锌版贴合印刷质量提高, 希望在贴合瞬间有一短暂的停留。

根据工艺动作分析, 简易式平版印刷机具有 3 个执行构件——油盘、油辊和印头。它们的运动形式分别为:

油盘作间歇转动, 一般采用在一个运动循环内作定向间歇转过  $60^\circ$  的动作。

油辊沿固定导路(它主要由油盘和铜锌版组成)在一个运动循环内作一次往复运动。

印头在一个运动循环内作一次往复摆动。

(3) 根据工艺动作顺序和协调要求拟定运动循环图

拟定简易平版印刷机运动循环图的目的是确定印头、油辊、油盘 3 个执行构件动作的先后顺序、相位, 以利对各执行机构进行设计、装配和调试。

在拟定运动循环图时要确定一个主要执行机构, 以它的主动件每转一周完成一个运动循环, 平版印刷机是以印头的执行机构的主动件的某一零位角为横坐标的起点, 纵坐标表示执行件的位移情况。在运动循环图上表示的位移曲线主要表达出运动的起迄位置, 而不必准确表示出各执行构件的运动规律。

图 47.5-31 所示为简易印刷机的运动循环图。印头的摆动具有工作行程和空回行程。油辊的摆动的工作行程是在印头回程中完成的。油盘在油辊工作行程后半段开始作间歇转动一次, 至油辊回程的前半段完成转动, 接着油盘停顿一直至第二次间歇运动开始。

拟定运动循环图时, 为了提高机器生产率, 可使各执行构件的动作起迄位置在不影响相互动作协调和干扰的前提下进行重叠安排。

确定了运动循环图后, 就可按此来拟定合适的运动规律, 进行机构设计。必要时, 可对所设计的机构进行运动分析, 用分析所得的位移规律加到初步设计的运动循环图上, 观察机构的运动是否协调, 评估机构的运动和动力性能是否合适。若有不当之处, 还可以将运动循环图作适当的修正。

#### (4) 机构选型

根据 3 个执行构件——印头、油盘、油辊的动作要求一般可以选择一些常用的、合适的机构。

对于印头执行机构, 一般可选择表 47.5-3 所示的 5 种机构。设计者也可根据需要另行构思和设计其他的机构。

对于油辊执行机构, 一般可选择如表 47.5-4 所示的 3 种机构: 曲柄摇杆机构加固定凸轮机构、摆动导杆机构加固定凸轮机构、六连杆机构加固定凸轮机构。

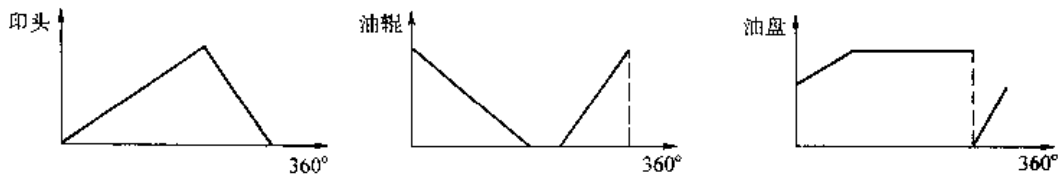


图 47.5-31 简易印刷机运动循环图

表 47.5-3 印头执行机构

序号	1	2	3	4	5
简图					
特点	结构简单、设计计算方便、有急回特性、全是铰链不易自锁	结构简单、设计计算方便、有急回特性、移动副中摩擦有影响	结构简单、设计计算方便、可有瞬时停歇、易磨损	结构比较复杂,可产生瞬时停歇、高副处易磨损	机构比较复杂、可产生瞬时停歇

表 47.5-4 油辊执行机构

序号	1	2	3
简图			
特点	结构简单、设计方便、但油辊刷墨速度不一定均匀	结构简单、设计方便、油辊刷墨速度难以均匀	结构比较复杂、设计也较难,但设法使油辊刷墨速度尽量均匀

对于油盘间歇运动机构,可以选择如表 47.5-5 所示的 4 种机构。在特殊情况下还可以采用利用连杆曲线的圆弧段或直线段来实现间歇运动。对于速度较低的平版印刷机一般可采用棘轮机构或槽轮机构。

表 47.5-5 油盘执行机构

序号	1	2	3	4
机构名称	棘轮机构	槽轮机构	不完全齿轮机构	凸轮式间歇运动机构
特点	结构简单,适用于低速,但需附加曲柄摇杆机构	结构简单,适用于低速,槽轮转角大小不能调节	结构比前两种机构复杂,具有瞬心线附加杆可减小冲击	凸轮形状复杂,制造较难,可用于高速场合

(5) 机械运动方案的选择与评定

从上述印头执行机构、油辊执行机构以及油盘间歇运动机构可以选择的种类数目考虑,在一般情况下,根据数学上排列组合原理,平版印刷机的机械运动方案数目有:

$$N = 5 \times 3 \times 4 = 60$$

从 60 种机构运动方案中,根据给定条件、各机构的相容性、要求机构尽可能简单等等来选择方案,如果印头不要求有瞬时停歇的保压阶段、油辊刷墨速度不考虑速度均匀,其机构运动方案有以下几种:

- 1) 曲柄摇杆机构-曲柄摇杆加固定凸轮机构-棘轮机构;

2) 曲柄摇杆机构-摆动导杆加固定凸轮机构-棘轮机构;

3) 摆动导杆机构-曲柄摇杆加固定凸轮机构-棘轮机构;

4) 摆动导杆机构-摆动导杆加固定凸轮机构-棘轮机构。

这4种方案,再加上间歇运动机构改为槽轮机构,也有4种方案,加起来一共有8种方案,从结构简单、摩擦情况良好考虑,在8种方案中可选用第一方案。

如果要求印头有瞬时停歇的保压阶段、油墨刷墨速度要考虑尽量均匀,在目前机器速度不高的情况下,各执行机构可选择如下:

1) 印头机构——采用表 47.5-3 中的第 4、5 两种机构;

2) 油辊机构——采用表 47.5-4 中的第 3 种机构;

3) 油盘机构——采用表 47.5-5 中的第 1、2 种机构。

因此,此时机械运动方案有 4 种,可以采用机械运动方案的评价方法(详见第 9 章)来评价择优。

(6) 机械传动系统的速比和变速机构

根据本例给定的条件,平版印刷机选用的驱动电机的转速为  $n=910\text{r/min}$ ,而印刷能力为  $24\text{次/min}$ (亦即平版印刷机的主轴转速为  $24\text{r/min}$ )。因此,必须采用减速机构,其减速比为:

$$i = \frac{n}{n_1} = \frac{910}{24} = 37.916 \approx 38$$

可采用一级带减速传动、二级齿轮传动,它们的传动比分别为:

1) 带传动:传动比为 3;

2) 第一级圆柱直齿轮传动:传动比为  $3.411, z_1=17, z_2=58$ ;

3) 第二级圆柱直齿轮传动:传动比为  $3.705, z_2'=17, z_3=63$ 。

(7) 画出机械运动方案简图(机械运动示意图)

根据上述确定的最简单的机械运动方案,画出机械运动示意图,如图 47.5-32 所示。其中包括由驱动电动机开始的机械传动系统,3 个执行机构组成的机械运动示意图。

(8) 对机械传动系统和执行机构的尺度设计

其内容包括有:

1) 对带传动进行初步设计计算;

2) 对第一对齿轮传动进行强度计算和几何尺寸计算,确定模数和有关尺寸;

3) 对第二对齿轮传动进行强度计算和几何尺寸计算,确定模数和有关尺寸;

4) 对印头执行机构-曲柄摇杆机构按摆角大小、

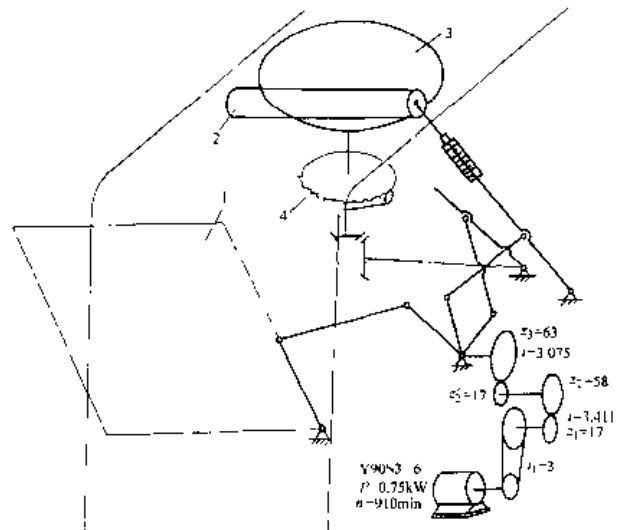


图 47.5-32 驱动电动机、传动机构示意图

行程速比系数等进行设计计算,必要时可进行机构运动分析后作改进设计;

5) 对油辊执行机构-曲柄摇杆机构按油辊摆动角度及机械运动循环图上的相关角关系进行设计计算,由于油辊执行机构是曲柄摇杆机构加固定凸轮机构组成,油辊的绝对运动求解是比较复杂的。上述组合机构相当于凸轮不是等速转动的移动从动件盘形凸轮机构。

6) 对油盘间歇运动机构-棘轮机构及其附加的曲柄摇杆机构的设计计算。

## 5.2 设计冲压式蜂窝煤成型机的运动系统方案

冲压式蜂窝煤成型机的运动系统设计主要步骤和过程如下:

(1) 冲压式蜂窝煤成型机的功能和设计要求

功能

冲压式蜂窝煤成型机是我国城镇蜂窝煤(通常又称为煤饼)生产厂的主要生产设备,这种设备由于具有结构合理、重量可靠、成型性能好、经久耐用、维修方便等优点而被广泛采用。

冲压式蜂窝煤成型机的功能是将粉饼加入转盘的模筒内,经冲头冲压成蜂窝煤。

为了实现蜂窝煤冲压成型,冲压式蜂窝煤成型机必须完成 5 个动作:

1) 粉煤加料;

2) 冲头将蜂窝煤压制成型;

3) 清除冲头和出煤盘内积屑的扫屑运动;

4) 将在模筒内冲压后的蜂窝煤脱模;

5) 将冲压成型的蜂窝煤输出。

### 设计要求和原始数据

1) 蜂窝煤成型机的生产能力: 30 次/min;

2) 图 47.5-33 表示冲头、脱模盘、扫屑刷、模筒转盘的相互位置情况。实际上冲头与脱模盘都与上下移动的滑梁连成一体, 当滑梁下冲时冲头将粉煤冲压成蜂窝煤、脱模盘将已压成的蜂窝煤脱模。在滑梁上升过程中扫屑刷将刷除冲头和脱模盘上粘着的粉煤。模筒转盘上均布了模筒, 转盘的间歇运动使加料后的模筒进入冲压位置, 成型后的模筒进入脱模位置, 空的模筒进入加料位置;

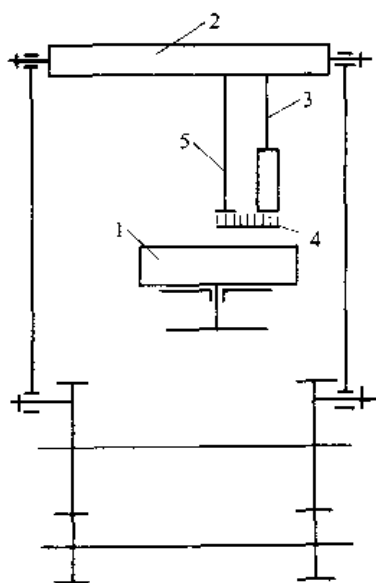


图 47.5-33 冲头、脱模盘、扫屑

刷、模筒转盘位置示意图

1—模筒转盘 2—滑梁 3—冲头

4—扫屑刷 5—脱模盘

3) 为了改善蜂窝煤冲压成型的质量, 希望冲压机构在冲压后有一保压时间;

4) 由于同时冲两只煤饼时的冲头压力较大, 最大可达 50 000N, 其压力变化近似认为在冲程的一半进入冲压, 压力呈线性变化, 由零值至最大值, 因此希望冲压机构具有增力功能, 以减小机器的速度波动, 减小原动机的功率;

5) 驱动电机目前采用 Y180L-8, 其功率  $P=11\text{kW}$ , 转速  $n=730\text{r/min}$ ;

6) 机械运动方案应力求简单。

#### (2) 工作原理和工艺动作分解

根据上述分析, 冲压式蜂窝煤成型机要求完成的工艺动作有以下 6 个动作:

1) 加料: 这一动作可利用粉煤重力打开料斗自动加料;

2) 冲压成型: 要求冲头上下往复移动, 在冲头行程的后 1/2 进行冲压成型;

3) 脱模: 要求脱模盘上下往复移动, 将已冲压成型的煤饼压下去而脱离模筒。一般可以将它与冲头一起固结在上下往复移动的滑梁上;

4) 扫屑: 要求在冲头、脱模盘向上移动过程中用扫屑刷将粉煤扫除;

5) 模筒转模间歇转动: 以完成冲压、脱模、加料 3 个工位的转换;

6) 输送: 将成型的煤饼脱模后落在输送带上送出成品, 以便装箱待用。

以上 6 个动作, 加料和输送的动作比较简单, 暂时不予考虑, 冲压和脱模可以用一个机构来完成。因此冲压式蜂窝煤成型机运动方案设计重点考虑冲压和脱模机构, 扫屑机构和模筒转盘的间歇转动机构这三个机构的选型和设计问题。

(3) 根据工艺动作顺序和协调要求拟定运动循环图

对于冲压式蜂窝煤成型机运动循环图主要是确定冲压和脱模盘、扫屑刷、模筒转盘三个执行构件的先后顺序、相位, 以利对各执行机构的设计、装配和调试。

冲压式蜂窝煤成型机的冲压机构为主机构, 以它的主动件的零位角为横坐标的起点, 纵坐标表示各执行构件的位移起迄位置。

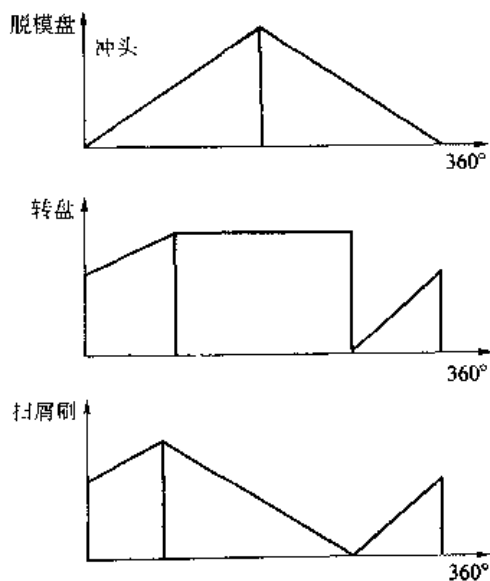


图 47.5-34 冲压式蜂窝煤成型机运动循环图

图 47.5-34 表示冲压式蜂窝煤成型机三个执行机构的运动循环图。冲头和脱模盘都由工作行程和回程两部分组成。模筒转盘的工作行程在冲头的回程后半段和工作行程的前半段完成。使间歇转动在冲压以前完成。扫屑刷要求在冲头回程后半段至工作行程前半段完成扫屑动作。

(4) 执行机构的选型

根据冲头和脱模盘、模筒转盘、扫屑刷这三个执行构件动作要求和结构特点,可以选择表 47.5-6 的常用

的机构,这一表格又可称为执行机构的形态学矩阵。

表 47.5-6 三执行机构的形态学矩阵

冲头和脱模盘机构	对心曲柄滑块机构	偏置曲柄滑块机构	六杆冲压机构
扫屑刷机构	附加滑块摇杆机构		固定移动凹轮移动从动件机构
模筒转盘间歇运动机构	槽轮机构	不完全齿轮机构	凸轮式间歇移动机构

图 47.5-35 表示附加滑块摇杆机构,利用滑梁的上下移动使摇杆 OB 上的扫屑刷摆动扫除冲头和脱模盘底上的粉煤层。图 47.5-35b 表示固定移动凸轮利用滑梁上下移动使带有扫屑刷的移动从动件顶出而扫除冲头和脱模盘底的粉煤层。

(7) 画机械运动系统方案简图

按已选定的三个执行机构的形式及机械传动系统,画出冲压式蜂窝煤成型机的机械运动系统示意图。其中三个执行机构部分也可以称为机械运动方案简图。如图 47.5-36 所示,其中包括了机械传动系统、三个执行机构的组合。如果再加上加料机构和输送机构,那就可以完整地表示整台机器的机械运动系统方案图。

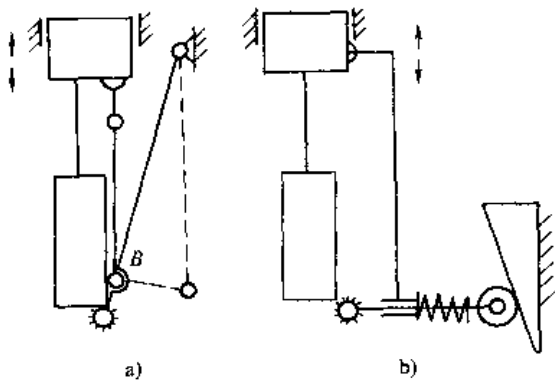


图 47.5-35 两种机构运动形式比较

(5) 机械运动系统方案的选择和评定

根据表 47.5-6 所示的三个执行机构形态学矩阵,可以求出冲压式蜂窝煤成型机的机械运动方案数为:

$$N = 3 \times 2 \times 3 = 18$$

现在,我们可以按给的条件、各机构的相容性和尽量使机构简单等等要求来选择方案。由此可选定两个结构比较简单的方案。

方案 I: 冲压机构为对心曲柄滑块机构,模筒转盘机构为槽轮机构,扫屑机构为固定凸轮移动从动件机构。

方案 II: 冲压机构为偏置曲柄滑块机构,模筒转盘机构为不完全齿轮机构,扫屑机构为附加滑块摇杆机构。

两个方案我们可以用模糊综合评价方法来进行评估选优,这里从略。最后选择方案 I 为冲压式蜂窝煤成型机的机械运动方案。

(6) 机械传动系统的速比和变速机构

根据选定的驱动电机的转速和冲压式蜂窝煤成型机的生产能力。它们的机械传动系统的总速比为:

$$i_{总} = \frac{n_{电机}}{n_{执行主轴}} = \frac{730}{30} = 24.333$$

机械传动系统的第一级采用带传动,其速比为 4.866;第二级采用直齿圆柱齿轮传动,其传动比为 5。

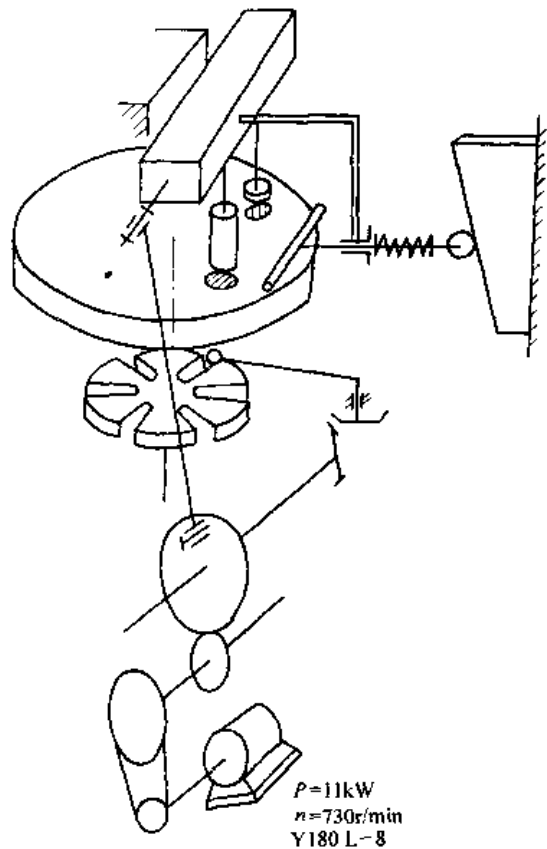


图 47.5-36 冲压式蜂窝煤成型机运动系统方案示意图

有了机械运动方案简图,就可以进行机构的运动尺度设计计算和机器的总体设计。

(8) 对机械传动系统和执行机构进行尺度计算

为了实现具体的运动要求,必须对带传动、齿轮传动、曲柄滑块机构(冲压机构)、槽轮机构(模筒转盘间歇运动机构)和扫屑凸轮机构进行运动学计算,必要时还要进行动力学计算。

**带传动计算**

1) 确定计算功率  $P_c$   $P_c = K_A P$

取  $K_A = 1.4$ , 则  $P_c = 1.4 \times 11 = 15.4 \text{ kW}$

2) 选择带的型号由  $P_c$  及主动轮转速  $n_1$ , 由有关线图选择 V 带型号为 C 型 V 带。

3) 确定带轮节圆直径  $d_1$  和  $d_2$  取  $d_1 = 200 \text{ mm}$ , 则

$$d_2 = 4.866 \times d_1 = 973.2 \text{ mm}$$

4) 确定中心距  $a_0$

$$0.7(d_1 + d_2) \leq a_0 \leq z(d_1 + d_2)$$

即  $821.24 \text{ mm} \leq a_0 \leq 2348.4 \text{ mm}$

5) 确定 V 带根数  $z$

$$z \geq \frac{P_c}{[P_0]} = \frac{15.4}{3.8} \approx 4$$

**齿轮传动计算**

取  $z_1 = 22$ ,  $z_2 = i \times 22 = 5 \times 22 = 110$ 。按钢制齿轮进行强度计算, 其模数  $m = 5 \text{ mm}$ 。则

$$d_1 = z_1 m = 110 \text{ mm}$$

$$d_2 = z_2 m = 550 \text{ mm}$$

其余尺寸, 按有关表格算出。

**曲柄滑块机构计算**

已知冲压式蜂窝煤成型机的滑梁行程  $s = 300 \text{ mm}$ , 连杆系数  $\lambda = \frac{R}{L} = 0.157$ , 则曲柄半径:

$$R = \frac{1}{2}s = 150 \text{ mm}$$

连杆长度  $L = \frac{R}{\lambda} = 955.41 \text{ mm}$

因此, 不难求出曲柄滑块机构中滑梁(滑块)的速度和加速度变化。

它的力分析也是比较容易的, 为简化起见, 不计各构件重量, 按冲压力变化作为滑块上受力。

**槽轮机构计算**

- 1) 槽数  $z$  按工位数要求选定为 6
- 2) 中心距  $a$  按结构情况确定  $a = 300 \text{ mm}$
- 3) 圆销半径  $r$  按结构情况确定  $r = 30 \text{ mm}$
- 4) 槽轮每次转位时主动件的转角  $2\alpha$   $2\alpha = 180^\circ$

$$\left(1 - \frac{2}{z}\right) = 120^\circ$$

5) 槽间角  $2\beta$   $2\beta = \frac{360^\circ}{z} = 60^\circ$

6) 主动件圆销中心半径  $R_1$   $R_1 = a \sin \beta = 150 \text{ mm}$

7)  $R_1$  与  $a$  的比值  $\lambda$   $\lambda = \frac{R_1}{a} = \sin \beta = 0.5$

8) 槽轮外圆半径  $R_2$   $R_2 = \sqrt{(a \cos \beta)^2 + r^2} = 262 \text{ mm}$

9) 槽轮槽深  $h$   $h \geq a(\lambda + \cos \beta - 1) + r$

$$h \geq 79.8 \text{ mm} \quad \text{取 } h = 80 \text{ mm}$$

10) 运动系数  $k$   $k = \frac{z-2}{2z} = \frac{1}{2}$  ( $n=1$ ,  $n$  为圆销数)

**扫屑凸轮机构计算**

固定凸轮采用斜面形状, 其上下方向的长度应大于滑梁的行程  $s$ , 其左右方向的高度应能使扫屑刷活动范围扫除粉煤。具体按结构情况来设计。

**(9) 冲压式蜂窝煤成型机的飞轮设计**

飞轮设计对于冲压式机械——三相交流电动机组成的机组可以采用精确算法, 为了简便, 我们采用了飞轮的近似算法, 其公式为:

$$J_M = \frac{[A]}{[\delta] \omega_m^2}$$

图 47.5-37 表示冲压式蜂窝煤成型机冲压力近似变化规律。假定驱动力为常值。则可求出  $P_d = 6250 \text{ N}$ 。

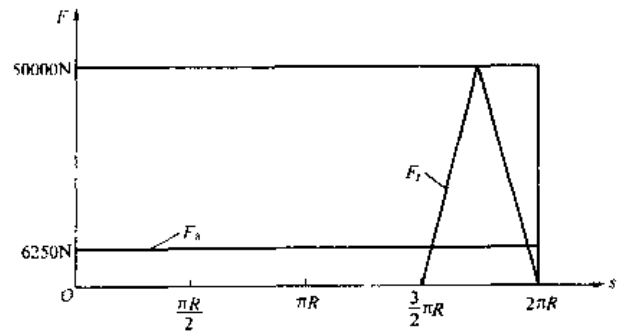


图 47.5-37 蜂窝煤成型机冲压力变化曲线

最大盈亏功  $[A]$ :

$$[A] = \frac{1}{2} \times (50000 - 6250) \times \frac{7}{8} \times \frac{\pi}{2} \times 0.15 = 4509.9 \text{ N} \cdot \text{m}$$

同时, 取  $[\delta]^* = 0.15$ , 为了减小飞轮尺寸, 将飞轮安装在小齿轮轴上, 则  $\omega_m = 150 \times \frac{2\pi}{60}$ , 因此,

$$J_M = \frac{4509.9}{0.15 \times \left(150 \times \frac{2\pi}{60}\right)^2} = 121.853 \text{ kg} \cdot \text{m}^2$$

加了飞轮之后, 由于飞轮能储存能量, 可使冲压式蜂窝煤成型机所需电动机功率减小, 其电动机功率约为

$$P' = Pdv = 6250 \times \frac{2\pi \times 0.15 \times 30}{60} = 2945.243 \text{ W} = 2.945 \text{ kW}$$

目前采用的电动机的功率为  $11 \text{ kW}$ , 显然没有考虑附加飞轮, 而是从克服短时冲压力较大的需要出发。



## 第6章 机电一体化系统设计

### 1 概述

#### 1.1 机电一体化系统的形成和发展

为了增强机械设备的功能和改善机械设备的性能,人们早在20世纪70年代以前就开始将机械与电子有机结合,产生了一些新颖的机械产品,为机电一体化产品的研制打下了基础。随着机械技术、微电子技术的飞速发展和广泛应用,人们日益注意在机械产品中,将机械技术与微电子技术相互集成和融合,使机电一体化技术得到迅速发展。正式提出“机电一体化”(mechatronics)概念是在1971年。因此,机电一体化技术可以认为已经历了30多年的发展。机电一体化技术正在不断完善,愈来愈受到重视。

20世纪70年代,日本学者首次提出了机电一体化的概念,当时只是将机械技术与电子技术简单结合,相应的机电一体化产品也比较简单,主要采用了高性能伺服技术,其典型产品如自动照相机、自动售货机等。20世纪80年代,由于高性能微处理器的出现并立即在机电一体化产品中得到应用,大大提高了机电一体化产品的自动化、智能化程度,使产品性能有较大提高,其典型产品如数控机床、工业机器人等。20世纪90年代,由于计算机网络和信息技术的飞速发展和广泛应用,使机电一体化产品向网络化、集成化发展。产品性能大大提高,具有远程操作等特性。通过网络远程控制进行心脏手术的机器人就是它的典型代表。

关于“机电一体化”(mechatronics)这个名词的起源,说法较多。早在1971年,日本“机械设计”杂志副刊就提出了“Mechatronics”这一名词,它是由Mechanics(机械学)与Electronics(电子学)组合而成的。从图47-6-1可见它是融合机械技术、电子技术、信息技术等多种技术为一体的新兴的技术。采用机电一体化技术设计和制造出的产品,称之为机电一体化产品。从系

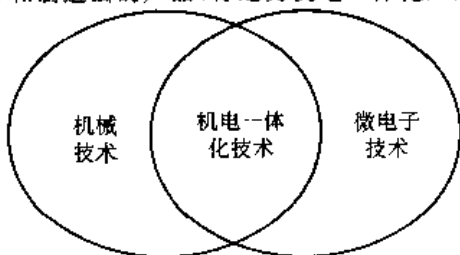


图 47-6-1 机电一体化技术

统科学的观点来看,机电一体化产品又可称之为机电一体化系统,它是集机械元件和电子元件于一体的复合系统。

#### 1.2 机电一体化系统的定义

机电一体化系统的英文译名“Mechtronics”最早是在1971年日本《机械设计》副刊特辑中提出,到了1976年前后已被日本各界所接受。但是对于机电一体化系统的定义,日本、美国、德国的学者认识并不一致。

在日本,“机械振兴协会经济研究所”于1981年3月提出:“机电一体化系统乃是机械的主功能、动力功能、信息功能和控制功能上引进微电子技术,并将机械装置与电子装置用相关软件有机结合而构成系统的总称。”

在美国,美国机械工程师协会于1984年也提出了对机电一体化系统的定义:“由计算机信息网络协调与控制的、用于完成包括机械力、运动和能量流等多动力学任务的机械和(或)机电部件相互联系现代机械系统。”

在德国,电气工程技术人员协会及其组成的精密工程技术专家组于1981年认为机电一体化系统是精密机械装置,“它包括机械(含液压、气动及微机械)、电工与电子、光学及其他不同技术的组合。”

从上述日本、美国、德国机械工程界对机电一体化系统的认识来看,我们认为机电一体化最本质的特征是一个机械系统。我们可将机电一体化系统定义如下:“机电一体化系统是将计算机技术融合于机械的信息处理和制功能,实现机械运动、动力传递和变换,完成设定的机械功能的现代机械系统。”

由于机械技术和微电子技术的有机结合,使现代机械系统更易实现功能的柔性化、智能化、自动化。

基于机电一体化系统是现代机械系统这一基本认识,将会有利于我们建立机电一体化系统的组成框架及其系统设计的过程模型。

#### 1.3 机电一体化系统的研究状况

总的来看,机电一体化系统设计的研究处于初始阶段。企业由于开发产品的需要,对于机电一体化系统设计理论和方法的需求十分迫切,但是现有的机电一体化系统的设计理论和方法远远无法满足这种需求。

目前,对于机电一体化系统设计的研究主要集中在

在欧洲、美国和日本。在欧洲开展机电一体化系统设计的研究的主要机构有：德国的 Darmstadt 大学、英国 Lancaster 大学的工程设计中心、荷兰 Twente 大学、比利时 Leuven 大学、挪威科技大学、丹麦技术大学和芬兰的 VTT 研究中心等。在美国开展机电一体化系统研究主要集中在 MIT(麻省理工学院)、Carnegie Mellon 大学、Michigan 大学和 Stanford 大学。在日本，东京大学的 Yoshikawa 和 Tomiyama 两位学者也在研究机电一体化系统设计。

德国 Darmstadt 大学的 R. Isermann, H. J. Herpel, M. Glesner 等人对机电一体化系统设计进行了较为深入的研究，集中研究了机电一体化系统的控制系统的设计方法学。他们认为控制系统设计是机电一体化的主体，较少考虑机械部分的特性、传感器的特性和驱动元件的特性，也没有考虑几部分之间的融合设计问题。

德国 Heinz Nixdorf 大学的 Jürgen Gausemeier, Martin Flath, Stefan Möhringer 等人于 2001 年构建了机电一体化系统开发的 V 型模型，指出在概念设计的早期阶段，需要有一种共同的功能描述语言来描述所涉及的不同学科的知识，给出了一种适用于机电一体化产品概念设计的集成方法，即用半规则式说明语言进行功能的功能原理建模。

在法国 PSA 所进行的 ESPFIT II/OLMECO (Open Library for Models of Mechatronics Components 机电一体化组成元件的模型的开放库)研究项目，取得了阶段性成果。该项目的核心是一个包含机电一体化基本功能元件，或者已在实践中应用的系统作为基本单元的模式库。这个模式库能够为工程设计人员提供正确的、有效的、可重用的模型单元。在机电一体化设计中可随时从库中调用这些基本元件或系统，从而提高了设计效率和设计水平。但所得方案往往缺乏创新性。

美国 Analogy 公司开发的 Saber 软件自称是为支持机电一体化系统智能化概念设计的，但实际最多是实现机电一体化系统的仿真。它只具备对机电一体化系统建模、性能仿真、灵敏度分析等功能，而不具备产生方案的功能。由于该软件是由电子系统设计软件发展而来，使其无法满足真正的机电一体化系统概念设计需要。

英国 Lancaster 大学的工程设计中心自 1990 年成立以来，在机电一体化系统概念设计方面作了大量的研究。他们采用键合图(Bond Graph)作为仿真技术的底层，通过能量守恒定律在各能量子系统之间搭起一座桥梁，采用方框图作为信息处理系统的底层表达方式。用面向功能模块的混合建模(包括键合图理论、方框图和高层次的功能模块表达)方法，构建了机电一体

化产品设计的 Schemebuilder 虚拟开发平台。通过系统模拟平台进行模拟、仿真、检验，从而使设计得到一定的验证并改善系统的性能。但该软件的设计偏重于控制部分，因此必然有它的局限性。

在我国，对机电一体化系统的概念设计问题还研究得不多。上海交通大学的邹慧君等在 1996 年开始对机电一体化系统概念设计问题作了较为系统的研究，并取得了较大的成果。他们针对目前国外对机电一体化系统概念设计较多均局限于电子软件和控制部分研究，认为机电一体化系统其本质上还是一个机械系统，是一个现代机械系统。这就是说机电一体化系统的主功能是机械功能，是为了实现运动和机械能的传递和变换。在机电一体化系统中实现信息的测试和传输、软件的编制和控制技术是围绕机械功能主线而进行的。通过机械、电子、软件的相互融合而使机械主功能实现自动化、智能化、柔性化和性能最优化。通过研究，提出了机电一体化系统组成的划分，机电一体化系统概念设计的设计模型及数学描述，基于知识重用的机电一体化系统方案组成，广义执行机构概念设计过程模型，广义执行机械的创新解法等等。

总之，机电一体化系统的设计理论和方法还有待深入研究。我们相信通过大家的努力，形成完整的机电一体化系统设计理论和方法，为期不会太远。

#### 1.4 机电一体化系统的组成

机电一体化系统设计过程模型的建立与机电一体化系统的组成部分密切相关。根据相关文献的阐述，对机电一体化系统的组成比较典型的有 3 种看法：

##### (1) 五块论

德国 Darmstadt 大学的 Rolf Isermann 提出机电一体化系统是由五大功能模块组成：控制功能、动力功能、传感检测功能、操作功能和结构功能。将机电一体化系统通俗地类比于人的大脑、内脏、五官、四肢及躯体。Rolf Isermann 认为控制功能模块是决定控制策略；动力功能模块是提供动力；传感检测功能模块是测量完成机械功能的相关信息；操作功能模块是指相应的执行元件，如电机等；结构功能模块是指机电一体化系统的壳体等。五块论从机电一体化系统方案设计阶段来看，存在一些缺陷：它没有强调从实现机械主功能出发去求取功能原理解；它过分强调了控制部分的作用；它忽视了实现运动传递和转换的执行机构系统，其实操作不仅仅是依靠电机等执行元件。如此等等，使我们对机电一体化系统缺乏深层次的认识，并影响我们深入解决机电一体化系统的设计问题。

##### (2) 三环论

丹麦理工大学的 Jacob Burr 等人提出机电一体化

系统好比是一个由机械、电子、软件三个相交圆环组成,认为机电一体化系统是由机械、电子、软件三大功能模块组成。其中机械模块包括执行机构、机械传动;电子模块包括驱动器的电力、电子部件和传感器;软件是指控制系统的软件。三环论的模型的特点是将功能按机械、电子、软件三个方面进行分配,以此表示机电一体化系统的组成和相互关联。机械、电子、软件的关联用三环来表达,相对比较含糊,令人难以琢磨,同时电子和软件的内涵也是没有明确的界定。由于上述两方面的模糊性使我们难以用此模型进行机电一体化系统方案设计。

### (3) 两子系统论

挪威科技大学的 Bassam A. Hussein 提出将机电一体化系统划分为两大子系统:物理系统与控制系统。物理系统包括各种驱动装置、执行机构、传感器等;控制系统包括控制系统的软、硬件。这种划分方式是从能量流和信息流来考虑的、能量流经的系统属物理系统,信息流经的系统属控制系统。两个子系统的划分从实现功能角度来看,比前两种组成的观点趋于合理,同时,有利于对控制系统的深入研究。但是,对物理系统的内涵不够细致,因为驱动装置和执行机构与传感器在实现功能方面不能统一。两个子系统论同样很难指导机电一体化系统的方案设计。

## 1.5 机电一体化系统组成的新认识

根据系统论的观点,将一个系统划分为互相联系的子系统的原则:一是应突出系统的主功能;二是应按功能分解原理以各相对独立的子功能来确定子系统。

从上述分析来看,机电一体化系统的主功能应是实现运动和动力的传递和转换,能量的传递和转换应服从运动和动力的传递和转换。机电一体化系统的执行机构系统就是为了实现机械的主功能。

从机电一体化系统的相对独立的子功能来看,实现机械的主功能所需参数的检测和传感的目的是实现相关参量信息的采集,为精确地实现控制提供信息。另外,实现信息处理及控制功能,使执行机构系统更好地完成运动的独立传递和转换。

从完成工艺动作过程这一总功能要求出发,机电一体化系统可以划分为:

1) 广义执行机构子系统——机电一体化系统的主功能,即运动及动力传递和转换功能是由此子系统来完成的。

广义执行机构由驱动元件和执行机构(或执行件)组成。如图 47.6-2 所示。在机电一体化系统中往往是由若干个广义执行机构组成广义执行机构子系统,它是完成运动及动力传递和转换的独立功能部分,应该

是现代机械系统(机电一体化系统)的主体。

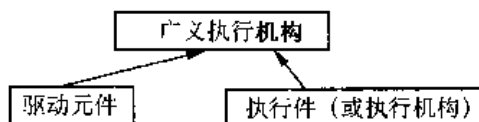


图 47.6-2 广义执行机构的组成

2) 传感检测子系统——机电一体化系统中用以检测机械参数和工作过程有关的参数的传感器、运算放大电路等,传感检测子系统的作用是采集控制广义执行机构所必需的参数信息,供信息处理及控制子系统进行处理并发出相关信息。这是控制广义执行机构子系统必不可少的环节。

3) 信息处理及控制子系统——机电一体化系统中将传感检测子系统取得的信息加以处理、运算从而发出控制广义执行机构子系统的控制信号。信息处理及控制子系统包括接口,微型计算机,系统整体控制方案及控制算法、软件设计等。信息处理及控制子系统的好坏,直接影响到整个机电一体化系统的性能和质量。这个子系统为机电一体化系统实现自动化、智能化、输出柔性化和性能优化提供了必要的条件。难怪不少人重视机电一体化系统的控制部分的研究,甚至把机电一体化系统的设计问题认为就是控制问题。其实脱开机械的主功能和机械的相关参数不可能设计出性能优良的机电一体化系统。一个性能优良的机电一体化系统应该将上述 3 个子系统很好地融合在一起。

图 47.6-3 表示机电一体化系统的子系统的组成和关联。提出机电一体化系统是由广义执行机构子系统、传感检测子系统和信息处理及控制子系统来组成,有利于明确机电一体化系统方案设计的目标、步骤和实施方法。因此,3 个子系统论是五块论、三环论、两个子系统论的发展和完善。将会有利于深化机电一体化系统设计年和方法的研究。

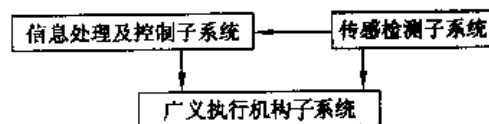


图 47.6-3 机电一体化系统的子系统的组成和关联

## 2 机电一体化系统的应用和特点

### 2.1 机电一体化系统的应用

机电一体化系统在机械产品中应用已十分广泛。机电一体化技术的应用使机械产品的结构、功用、性能等方面产生了重大的变化。具体来说表现在如下几个方面:

(1) 机电一体化技术可以实现机械产品的高性能

和多功能

在原有机械产品中引进了机电一体化技术后,使其自动化程度提高,工艺动作易丁复杂多变,工艺质量提高和能耗降低等等。例如,对于工艺动作复杂,产品更新换代快的轻工机械、纺织机械、印刷机械、包装机械和制药机械等,所属行业都十分重视产品的机电一体化。机械产品的机电一体化使产品的技术含量和附加值大为提高,有利于适应日益剧烈的市场竞争需要。

(2) 用电子控制器件取代机械中一部分机械控制机构

用电子控制器件替代机械控制机构可以扩大原有机械产品的功能,提高工作性能。如普通机床加上数控系统,普通缝纫机上增加电子控制的绷架系统都可以使机械产品水平大为提高。

(3) 以电子器件为主的机电共存的产品

这类产品一般来说机械结构比较简单,但要考虑机电两者的融合以提高产品的性能。复印机、录音机、录像机等等均属于这类产品。

从以上情况来看,机电一体化系统是有不同层次和应用场合的。其机电融合的程度也是因具体产品而异。

## 2.2 机电一体化产品的主要特点

机电一体化产品与传统的机械产品相比,有如下的一些显著特点:

(1) 机械结构简单化

一台传统的机械设备往往需要采用比较复杂的机械传动系统来连接各个相关的执行构件,以保证各执行构件动作的同步性和协调性。采用机电一体化技术后,可以改用几台电动机分别驱动,用电子器件、微型计算机来控制 and 实现各执行构件的动作,完成机械的工艺动作过程。例如:一台微机控制的精密插齿机,可以节省齿轮等传动部件约占30%;又如用单片机控制针脚花样的电脑缝纫机,可以比老式缝纫机减少了350个机械零部件。

(2) 提高了加工工艺的精度

由于机械传动部件的减少,机械磨损及间隙配合等所引起的运动误差大为减少,通过微机控制系统可以精确地按照预先给定量,同时可使相应的机械动作、各种干扰因素造成的误差,进行自行校正、补偿,从而可以达到单纯机械方法所实现不了的加工工艺精度。例如:微机控制的精密插齿机加工的圆柱齿轮的精度,可以比原有的插齿机高一个精度等级。又如大型镗、铣床,只要安装上感应同步器数显装置,即可使加工精度从0.06mm/1000mm提高到0.02mm/1000mm。

(3) 机械工艺动作过程的柔性化

由于机电一体化系统中广义执行机构子系统广泛采用微机控制。因此,只要改变控制的计算机程序,就能改变设备的加工能力和机械工艺动作过程,以迅速适应改变被加工产品的结构、满足多品种、小批量的需要,使加工生产线更具柔性化。

(4) 工艺动作过程的智能化

由于机电一体化系统采用了传感检测子系统和信息处理及控制子系统,因此可以按人工智能要求确定工艺动作,实现机械的智能化。例如自动调焦和调光圈的自动照相机,又称为“傻瓜照相机”。实际上它是一台智能化的照相机。

(5) 操作自动化

采用传感检测子系统和信息处理及控制子系统后,可以依靠微机实现一台机器各个相关传动机构的动作及它们的功能协调关系,实现机械产品操作的全部自动化。例如:一般的数控机床加工零件时,将被加工零件的工艺过程、工艺参数、机床运动要求等,用数控语言记录在数控介质(如磁盘)上,然后输入到机床的数控装置,再由数控装置控制机床工艺动作过程,从而实现加工自动化。

(6) 调整维修方便

机电一体化产品在使用现场安装调试时,一般均可通过控制程序的变更来实现工作方式、工艺动作过程的变动,以适应各个用户工作对象及现场工作参数变化的需要。同时,机电一体化产品的维修也比较方便。

## 3 机电一体化系统设计过程模型及数学描述

机电一体化系统设计过程模型是建立在系统的功能需求和组成部分上的。功能需求是系统设计的出发点和归宿,组成部分是实现系统目标的主要基础。采用3个子系统(即广义执行机构子系统、传感检测子系统、信息处理及控制子系统)作为机电一体化系统的组成部分,实际是从完成机械主功能、实现工作过程参数测量以及完成过程控制这3个子功能出发的。通过将3个子功能及其行为、结构的集成就可以实现机电一体化系统方案的设计。

### 3.1 机电一体化系统设计过程模型的建立

机电一体化系统设计过程模型的建立要充分考虑计算机对机电一体化系统方案设计的支持。

图47.6-4所示为完成某一动作的广义执行机构的设计过程。从设计过程来看,它是构成了功能-行为-结构的设计模型。而某一机械分功能与检测功能、控制功能融为一体。机械分功能又是三者的主体。

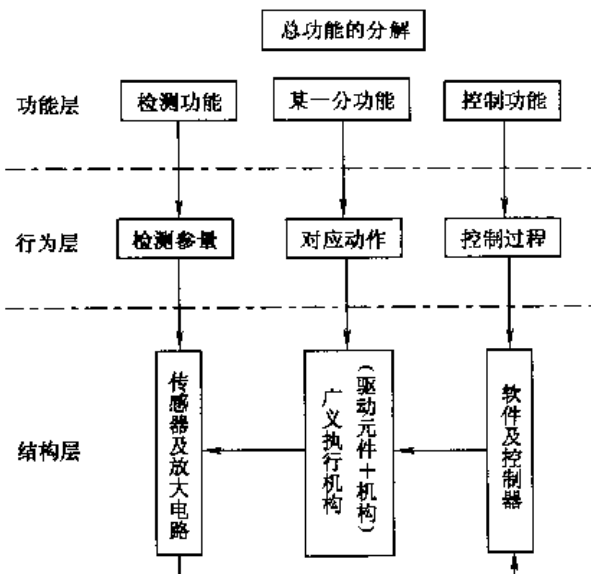


图 47.6.4 单一动作广义执行机构的设计过程

图 47.6-5 表示机电一体化系统方案设计过程模型,具体表述了系统的设计过程和广义执行机构子系统、检测传感系统、信息处理及控制子系统之间的关联。

很显然在机电一体化系统组成之前,应该先确定围绕一个广义执行机构的检测-控制部分的一个子集成。如图 47.6-4 所示,进行此集成的构思和设计,产生若干个可行的方案,作为机电一体化系统方案设计时的重要依据。

从图 47.6-5 所示的机电一体化系统方案设计过程来看,检测传感相应于系统的各个分系统、信息处理及控制子系统的各个分系统均与相应的广义执行机构子系统的各个分系统紧密融合在一起,形成了 3 个子系统之间的联系。

为了实现机电一体化系统方案综合最优的选择,我们必须进行评价与决策。对于机电一体化系统方案

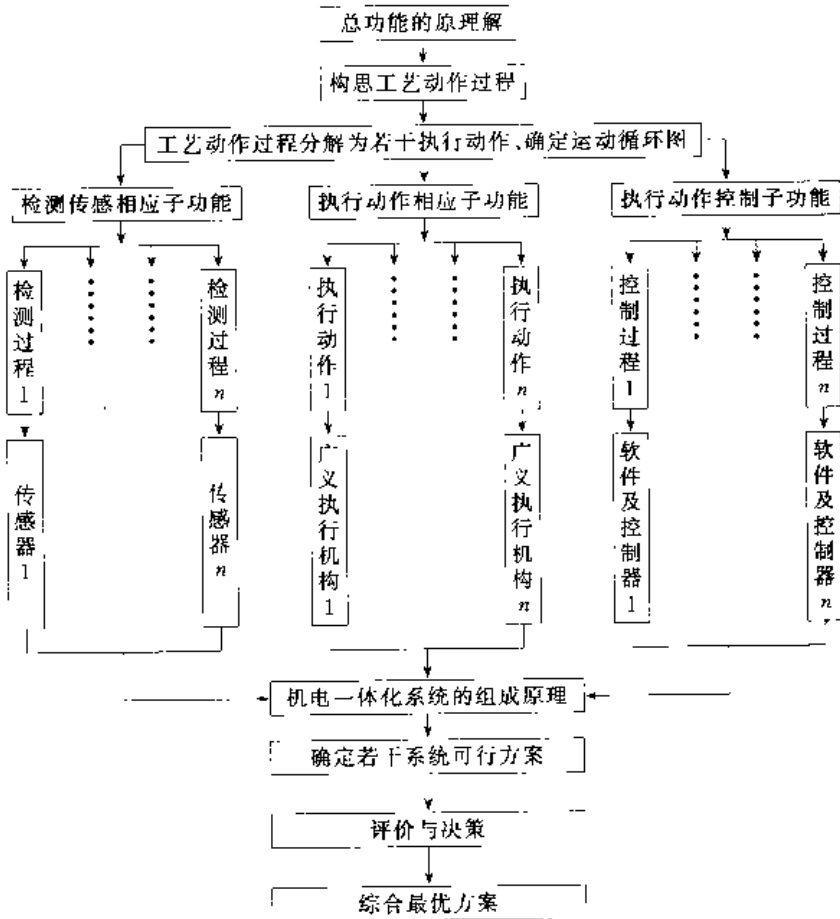


图 47.6-5 机电一体化系统方案设计过程模型

的评价可以采用功能—行为—结构三层决策模型如图 47.6-6 所示。具体来说,先对每个子集成—执行—检测—控制子集成从功能、行为、结构三个层次来加以评价分析,然后对整个系统从功能、行为、结构三个层次上来进行评价分析,最后进行决策。

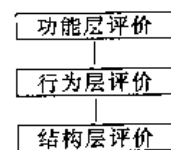


图 47.6-6 三层决策模型

### 3.2 机电一体化系统设计过程的数学描述

机电一体化系统方案设计是一个基于知识的求解过程,它始于设计要求,止于获得系统方案。因此机电一体化系统方案设计的数学描述系统应是遵循功能—行为—结构的思维过程。

机电一体化系统方案设计的进程应表达为图 47.6-7 所示的步骤。

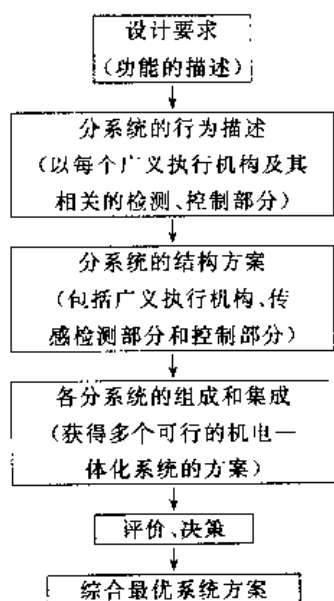


图 47.6-7 机电一体化系统方案设计进程

对于机电一体化系统从组成来看,它是由三大子系统(即广义执行机构子系统、传感检测子系统和信息处理及控制子系统)来构成的,这是一种横向结构型式。但是从设计进程来看,它是由各分系统(以每个广义执行机构及相关的检测、控制部分所构成的)行为描述和结构方案确定;最后将各分系统集成成为系统方案。这是一种纵向设计进程。

#### (1) 分系统方案设计的数学描述

分系统方案设计由两个阶段组成——方案的综合和方案的评价。

方案的综合过程主要在设计要求的基础上,通过行为的描述,提出分系统方案的备择集。从设计要求到方案备择集的映射过程可以定义如下:

$$S_i = K^s(R_i^d, B_i^d) \quad (47.6-1)$$

式中  $S_i$ ——分系统的备择方案;

$R_i^d$ ——设计要求,应包括广义执行机构、检测部分、控制部分要求;

$B_i^d$ ——行为描述,应包括广义执行机构、检测和部分的描述;

$K^s$ ——可用知识域(包括机械、传感、控制、软件等技术)。

方案评价过程是以分系统评价指标集为依据对备择方案集进行评估,其决策集可以计算如下:

$$B_i = A_i \cdot R_i \quad (47.6-2)$$

式中  $B_i$ ——分系统的决策集;

$A_i$ ——分系统的权数分配集;

$R_i$ ——分系统的评价矩阵。

以此来计算出方案备择集中各方案的好坏。

#### (2) 机电一体化系统方案设计的数学描述

根据上述评价方法可以选择每种分系统的若干优秀的方案,以此组合成该机电一体化系统的备择集。它们的评价过程中的决策集,可以作如下计算:

$$B = A \cdot R \quad (47.6-3)$$

式中  $B$ ——系统的决策集;

$A$ ——系统的权数分配集;

$R$ ——系统的评价矩阵。

比较各方案的决策集,不难求得综合最优的系统方案。

## 4 广义执行机构子系统的类型和设计

### 4.1 传统执行机构

在机器中执行机构是用作运动和动力的传递和变换。

18世纪下半叶工业革命推动了生产的机械化,产生了纺织机械、蒸汽机以及内燃机等等。以机械化为主要特征的传统执行机构,就是这一时代的产物。那时对机构定义为:具有确定运动的构件系统。构件均看作刚性杆件,且认为运动副中无间隙。它的结构学、运动学和动力学目前已比较完备。

在传统的机械系统中往往是由一个驱动元件(如电动机、液压马达等)等速驱动各执行机构,依靠各执行机构输入件的相位角布置和传动机构,使各执行机构进行协调动作完成机械的工艺动作过程。因此,传统执行机构只要研究它的类型和运动学、动力学,不必与驱动元件的特性联系在一起进行研究。这种传统的执行机构又可称之为无源机构,其输出运动取决于机构的类型和机构的运动尺寸。在构件为刚体、副中无间隙的假定条件下,机构看作是几何图形。因此,将机构的运动学称之为运动几何学,使机构学的研究相对比较单纯。

### 4.2 广义执行机构

20世纪70年代以来,随着科学技术的发展,机械技术与电子技术、控制技术、信息技术、传感技术等相互结合,产生了新一代的机械系统——机电一体化系统,这种现代机械系统的主要特征是可控性,其中各

执行机构的输出运动都是可控的。广义机构的概念由此而产生。

广义机构与传统机构主要区别有3点:

1) 广义机构有别于传统机构,是引入了柔性件、弹性件等非刚性构件,使机构脱离了纯刚性的范畴。机构同时更具可变性和复杂性;

2) 广义机构有别于传统机构,是由于融合了控制技术,使驱动元件的输出机械特性更加符合机构多变的输出需要;

3) 广义机构有别于传统机构,是将驱动元件与机构集成为一体,使机构的输出更具真实性和可控性。

根据上述讨论,我们对广义机构定义如下:广义机构是实现可控运动或不可控运动的驱动元件与由刚性、非刚性构件组成的运动链两者集成为一体的系统。

这里所指的驱动元件的种类是十分广泛的,有感应电动机、步进电机、伺服电机、液压缸、气动缸、弹簧、重锤、形状记忆合金、光电马达……

驱动元件的种类如图 47.6-8 所示。

在广义机构中由于驱动元件与构件系统融为一体,使输出运动可以复杂多变,使广义执行机构具有可控性。即使机构各构件仍为刚性,其输出运动复杂多变的特性仍然具备,适应了机电一体化系统设计的需要。

另外,如果执行构件(输出构件)直接与驱动元件相连,不能说是机构的消失,而从传统机构学来看,它是一个由运动构件与机架相构成的二杆机构。因此,无疑它还归属于执行机构。

不可控电机和由刚性构件组成的机构的集成系统,与传统机构在形式和组成上已是不同的。我们只是将它看成广义机构中比较特殊的一种类型。

总之,广义机构概念的形成,是与机电一体化系统划分成机械主功能、检测功能、控制功能密切相关的。广义机构概念的提出是有利于进行机电一体化系统方案的构思和设计。

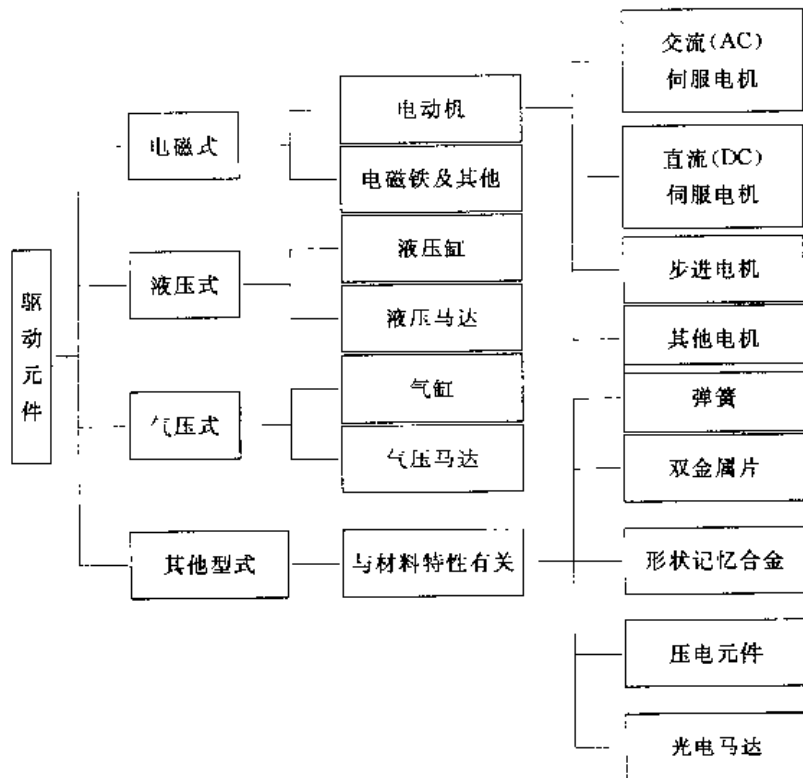


图 47.6-8 驱动元件的种类

### 4.3 广义执行机构的种类和基本特性

(1) 广义执行机构的种类

按驱动元件的类型分类

1) 电动型广义机构:驱动元件采用各种电动机的广义机构;

2) 液压、气动型广义机构:驱动元件采用各种液压、气动驱动元件的广义机构;

3) 弹性元件型广义机构:驱动元件为弹簧、簧片等的广义机构;

4) 形状记忆合金型广义机构:采用形状记忆合金为驱动元件的广义机构;

5) 电磁型广义机构:采用电磁元件驱动的广义机构;

6) 压电元件型广义机构:采用压电晶体等作为驱动元件的广义机构;

7) 其他驱动型式的广义机构。

按被驱动的机构类型分类

1) 两杆机构及多杆机构:驱动元件直接驱动执行件的为两杆机构,其余多杆机构包括凸轮机构、齿轮机构、连杆机构等等;

2) 单自由度机构及多自由度机构:被驱动的机构自由度可以是单自由度也可以是多自由度。多自由度机构的驱动元件数应与自由度一致;

3) 弹性机构和刚性机构:被驱动机构的构件全部是刚性的称刚性机构,而被驱动机构的构件中至少有一为弹性的称之为弹性机构;

4) 控制方式不同的可控机构:如采用调节构件运动尺寸的可调机构、采用编程控制的可编程机构、信息反馈的自动控制机构、混合输入的伺服控制机构等;

5) 开链机构及闭链机构:被驱动的是开链称开链机构,如开链机器人。被驱动的是闭链称闭链机构。

(2) 广义机构的基本特性

广义机构由于是驱动元件和机构的集成体,它的基本特性如下:

1) 机电一体化 广义机构由于其具有可控性,易于实现机电一体化,使其在现代机械中广泛采用。通过传感技术、电子技术、控制技术使机电融合于一体,实现机构的现代化。

2) 输出运动的多样化 通过对驱动元件的可编程控制,使机构的输入运动按需要改变,得出多种多样满足生产需要的输出运动。

3) 输出运动间歇停顿可能性 普通的四杆机构难以实现间歇停顿,但是通过对驱动元件控制实现任意停顿,使机构的运动与动力性能大为改善。

4) 智能化 通过采用一些智能化驱动元件,如形状记忆合金,使机构输出运动智能化,实现机器的智能控制。

5) 微型化 通过微型马达、压电晶体等作用可使机构产生微米级工作行程,实现机构微型化。

### 4.4 驱动元件的机械特性和基本特点

驱动元件的机械特性是指驱动力(或力矩)与运动参数(如转速、转角等)的关系。

(1) 直流伺服电机

直流伺服电机的功能是将输入的电压控制信号快速转换为轴上的角速度和角位移输出、直流伺服电机具有如下特点:

1) 可控性好:具有线性的调节特性,其转速正比于控制电压的大小,转向取决于控制电压的极性(或相位);控制电压为零时,转子能立即停转。

2) 稳定性好:能在转簧的转速范围内稳定运行。

3) 响应快:具有较大的起动转矩和较小的转动惯量,在控制信号发生变化时,伺服电机的转速能快速跟随变化。

直流伺服电动机广泛应用于精确位置控制系统和宽调速系统中。

在控制电压一定时,输出转矩  $T$  与转速  $n$  的关系,如图 47.6.9 所示。

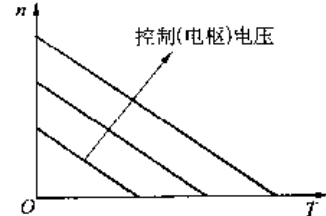


图 47.6-9 直流伺服电机机械特性

(2) 交流伺服电机

笼型交流伺服电机的主要性能为:励磁电流较小,体积较小,机械强度较高,低速运转时不够平滑,有抖动现象。它的使用范围为小功率自动控制系统、随动系统和计算装置中心。它的机械特性如图 47.6-10 所示。它的机械特性曲线是非线性的。非线性使系统产生动态误差。

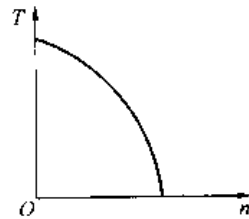


图 47.6-10 交流伺服电机机械特性

(3) 步进电机

步进电机是一种把电脉冲信号转变成角位移或直线位移的驱动元件。电脉冲由专用电源供给。每输入一个脉冲,电动机前进一步,故又称为脉冲电机。步进电机的位移量与脉冲数成正比,速度与脉冲频率成正比;在其负载能力范围内,不因电源电压、负载、环境条件的波动而变化;可以在宽广的范围内通过改变脉冲频率来调速;能够快速起动、反转和制动;在一相绕组长期通电时具有自锁能力。

表 47.6-1 表示常用步进电机分类和特点。

图 47.6-11 表示步进电机与驱动电路框图。其中  $\theta$  为输出转角,  $N$  为输入脉冲数。

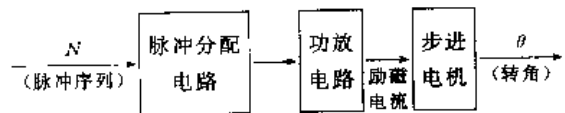


图 47.6-11 步进电机与驱动电路框图

图 47.6-12 表示步进电机的机械特性曲线。

图 47.6-13 表示步进电机如何选取脉冲速率。在此临界转矩状态下使用时,其运动方程式为:



表 47-6-1 常用步进电机分类和特点

序号	名称	特点
1	反应式步进电机	起动和运行频率较高,断电时无定位转矩,消耗功率较大
2	永磁式步进电机	消耗功率比反应式步进电机小,但需供给正、负脉冲电流,起动和运行频率较低,有定位转矩
3	永磁感应子式步进电机	有较高的运行和起动频率,需正、负脉冲供电,消耗功率较小,有定位转矩,具有反应式和永磁式步进电机两者优点
4	直线步进电机	可直接提供直线运动,系统结构简化,惯量小,提高了系统的快速性和精度,显著改善了系统的动态性能

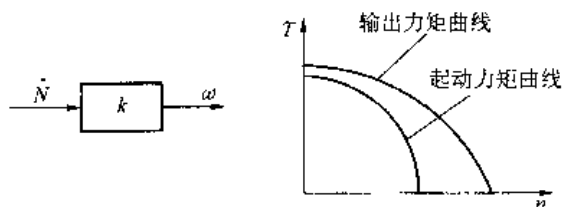


图 47.6-12 步进电机的机械特性  
 $J_M \omega = T_0 - K_E \omega$  (47.6-4)

其中,  $J_M$  为惯性负载;

$K_E$  为电枢电势常数。

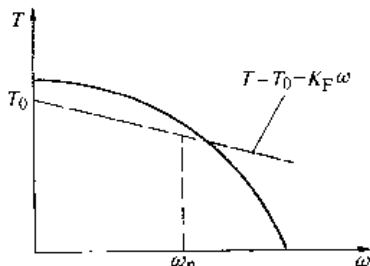


图 47.6-13 脉冲速率的选取

(4) 电磁铁

电磁铁是利用电磁吸力来操纵或者牵引机械装置或执行机构以完成自动化的动作。

图 47.6-14 为电磁铁驱动元件示意图。图 47.6-15 表示电磁铁驱动的机械特性曲线。

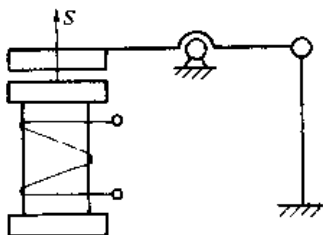


图 47.6-14 电磁铁驱动元件



图 47.6-15 电磁铁驱动的机械特性

(5) 弹簧

弹簧是利用弹性变形所产生的弹性力来驱动执行

机构以完成某种执行动作。弹簧一般又为片簧、螺旋弹簧以及盘簧等等。它们的弹性力表达式为  $P=KS$ 。

图 47.6-16 表示弹簧驱动的机械特性曲线。

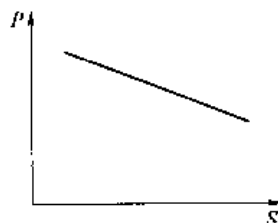


图 47.6-16 弹簧驱动的机械特性

(6) 压电式驱动元件

压电式元件具有给予应力应变时,则产生电荷;反之给予压力,则产生应力应变的性质。后一性能就是压电式驱动元件的工作原理。

图 47.6-17 为压电式驱动元件的驱动特性曲线。压电驱动元件用于产生微小位移,其优点是没有机械滑动。但压电元件有磁滞现象。

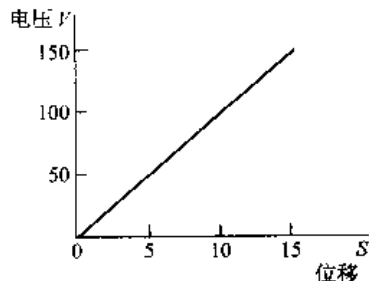


图 47.6-17 压电元件特性( $\mu\text{m}$ )

(7) 形状记忆合金

利用 Ti-Ni 或 Cu-Al-Ni 的合金材料冷却至马氏体相变终了温度  $M_f$  以下,并施加应力使产生塑性变形;若重新加热至  $M_f$  以上,合金因返回原取向,马氏体发生逆转,形状得以恢复,这种现象称形状记忆效应。形状记忆合金可用作产生微动的驱动元件。

(8) 微电动机

目前已有五类微电动机,即静电电动机、超声电动机、电磁电动机、谐振电动机和生物电动机,其中静电电动机应用最多。

静压电动机利用两充电电极之间基于静电能的能量变化趋势可产生机械位移。目前,已经在硅材料上制作出了多种类型的直径为  $30\sim 100\mu\text{m}$  的转动电动机,最快的电动机转速超过了  $4\ 000\text{r}/\text{min}$ 。

#### 4.5 驱动元件与执行机构的匹配

广义执行机构是由驱动元件与执行机构集合而成,为了使广义执行机构工作性能优良,应考虑驱动元件与执行机构在机械特性、工作精度、运转速度及能量等的匹配。

##### (1) 驱动元件和执行机构负载的机械特性匹配

广义执行机构中驱动元件的机械特性应与执行机构负载机械特性相匹配。使广义执行机构的速度和负载变化时,驱动元件的驱动力的机械特性,即驱动力矩与转速(或转角)的变化曲线能适应。这种驱动元件机械特性与执行机构负载机械特性的适应性就是机械特性的匹配。

##### (2) 驱动元件和执行机构的工作精度匹配

广义执行机构的工作精度取决于执行机构输出的精度,这种执行机构输出精度取决于机构的类型、传动精度。但是广义执行机构的输出精度还与驱动元件的工作精度有关。因此,为了保证广义执行机构输出精度,必须选择合适的驱动元件的类型。

##### (3) 驱动元件与执行机构运转速度的匹配

在广义执行机构中考虑到驱动元件和执行机构负载的机械特性匹配、工作精度最佳以及其他的工作要求,驱动元件的最佳工作转速与执行机构的运转速度不是相同的。为了使运转性能优良,往往需在两者之间增加一个传动机构,使驱动元件与执行机构的运转速度能相互匹配,提高广义执行机构的效能。这种传动机构也可认为包容在执行机构之内。因此,广义执行机构可以将传动机构与执行机构统称为执行机构,这在不少机电一体化系统中就是如此对待。

##### (4) 驱动元件与执行机构负载能量的匹配

广义执行机构负载的功率计算是按执行机构在一个工作循环中负载所作功来决定的。广义执行机构中驱动元件的驱动功率应按下式计算:

$$P_{\text{驱动}} = \frac{P_{\text{负载}}}{\eta_{\text{总}}} \quad (47.6-5)$$

式中  $P_{\text{驱动}}$ ——驱动元件的输入功率;

$P_{\text{负载}}$ ——执行机构负载的功率;

$\eta_{\text{总}}$ ——广义执行机构的总效率。

#### 4.6 广义执行机构的运动方程式

对于单自由度广义执行机构,我们一般把它简化成一个等效构件来研究。等效构件的运动方程式一般有两种形式:

##### (1) 能量方程式

对力矩形式为

$$\int_{\varphi_1}^{\varphi_2} M_d d\varphi - \int_{\varphi_1}^{\varphi_2} M_r d\varphi = \frac{1}{2} J_2 \omega_2^2 - \frac{1}{2} J_1 \omega_1^2 \quad (47.6-6)$$

式中  $\varphi_1, \varphi_2$ ——等效构件在 1, 2 位置处转角;

$M_d, M_r$ ——分别为等效驱动力矩, 等效负载力矩;

$J_1, J_2$ ——广义执行机构在转角  $\varphi_1, \varphi_2$  处的等效转动惯量;

$\omega_1, \omega_2$ ——广义执行机构在转角  $\varphi_1, \varphi_2$  处的角速度。

对力形式为:

$$\int_{s_1}^{s_2} P_d ds - \int_{s_1}^{s_2} P_r ds = \frac{1}{2} m_2 v_2^2 - \frac{1}{2} m_1 v_1^2 \quad (47.6-7)$$

式中各符号的含义,与上式相类似。

##### (2) 力矩方程式

由  $(M_d - M_r) d\varphi = \frac{1}{2} d(J\omega^2)$

$$\text{得出: } M_d - M_r = J \frac{d\omega}{dt} + \frac{dJ}{d\varphi} \frac{\omega^2}{2} \quad (47.6-8)$$

相似地可写成:

$$P_d - P_r = m \frac{dv}{dt} + \frac{dm}{ds} \frac{v^2}{2} \quad (47.6-9)$$

由运动方程式按数值解法,求得广义执行机构的真实运动。

## 5 检测传感子系统类型和设计

检测传感系统主要包括传感器和微机的接口两部分。

### 5.1 检测传感器的分类与基本要求

传感器是将机电一体化系统中被检测对象的各种物理变化量变为电信号的一种参量变换器。它主要功用是检测机电一体化系统自身与作业对象、作业环境的状态,为有效地控制机电一体化系统的动作提供信息。

#### (1) 检测传感器的分类

传感器的种类繁多,分类方法也有多种。

按作用可分为检测机电一体化系统内部状态的内部信息传感器和检测作业对象和外部环境状态的外部信息传感器。具体见表 47.6-2。

按输出信号的性质,可将传感器分为开关型(二值型)、模拟型和数字型三种,具体见表 47.6-3。

#### (2) 检测传感器的基本要求

机电一体化系统对检测传感器的基本要求有:

- 1) 体积小、重量轻,对整体的适应性好;
- 2) 精度和灵敏度高,响应快,稳定性好,信噪比高;

表 47.6-2 按作用分为内部信息传感器和外部信息传感器

名称		传感器具体类型
内部信息传感器		位置传感器、速度传感器、力传感器、力矩传感器、温度传感器
外部信息传感器	与人体五官对应的	接触式
		非接触式
	纯工程性的	视觉传感器、听觉传感器
		电涡流传感器、超声波测距仪、激光测距机

表 47.6-3 按输出信号的性质对传感器分类

名称	工作原理	传感器具体类型
开关型 (二值型)	接触式	微动开关、接触开关
	非接触式	光电开关、接近开关
模拟型	电阻型	电位器、电阻应变片等
	电压、电流型	热电偶、光电电池、压电元件等
	电感、电容型	电感式位移传感器、电容式位移传感器等
数字型	计数型	二值计数器
	代码型	编码器、磁尺

- 3) 安全可靠、寿命长;
- 4) 便于与计算机连接;
- 5) 不易受被测对象特性(如电阻、导磁率)的影响,也不影响外部环境;
- 6) 对环境条件适应能力强;
- 7) 现场处理简单,操作性能好;
- 8) 价格便宜。

(3) 开关型、模拟型和数字型传感器的工作原理

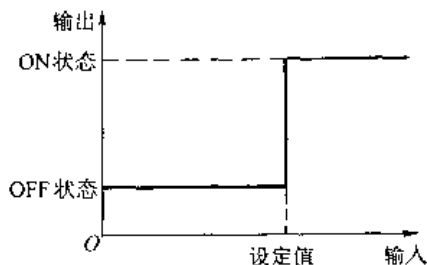


图 47.6-18 二值型传感器的工作原理

1) 开关型传感器 开关型传感器又称二值型传感器。它的二值就是“1”和“0”或开(ON)和关(OFF)。其工作原理见图 47.6-18 所示。这种“1”和“0”数字信号可直接传送到微机进行处理,使用方便。

2) 模拟型传感器 模拟型传感器的输出是与输入物理量变化相对应的连续变化的电量。

图 47.6-19 表示传感器的输入-输出关系,图 a 为线性的,图 b 为非线性的。

线性输出信号可直接被采用,而非线性输出信号则需进行适当修正,将其变成线性信号。一般需先将这些线性信号进行模-数(A/D)转换,将其转换成数字信号再送给微机进行处理。

3) 数字型传感器 数字型传感器分为计数型和

代码型两大类。

计数型又称脉冲数字型,其工作原理如图 47.6-20 所示,它是一种脉冲发生器,所发出的脉冲数与输入量成正比,并用计数器计数。增量式光电码盘就是如此。

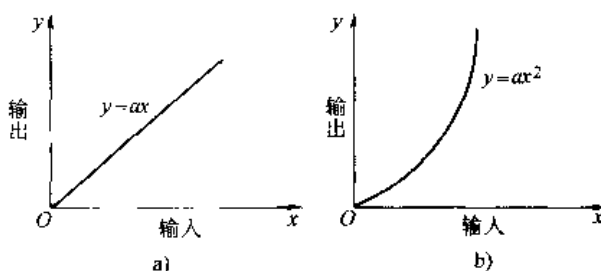


图 47.6-19 模拟型传感器的工作原理

代码型传感器又称编码器,其工作原理如图 47.6-21 所示,它输出的信号是数字代码,每一代码相当于一个一定的输入量之值。若高电平代码为“1”,低电平代码为“0”。从四条脉冲输出看,输入量为  $K_1$  时,输出代码为 1010。绝对值型光电编码器就是如此。

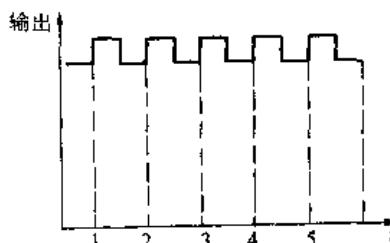


图 47.6-20 计数型传感器工作原理

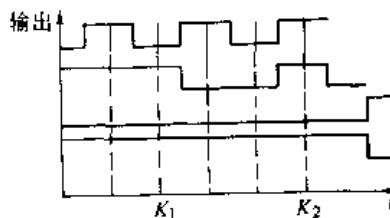


图 47.6-21 代码型传感器工作原理

## 5.2 位移检测传感器

位移检测传感器用来直接或间接检测目标的位移。表 47.6-4 表示了常用的位移检测传感器的种类、工作原理和特点。

表 47.6-4 位移检测传感器的分类、工作原理及特点

运动型式	传感器类型	传感器名称	工作原理	特点
直线型	光电型	光电编码器	是一种脉冲发生器,用数字代码检测线位移	结构较简单,使用方便,精度高
		光栅	利用光栅的莫尔条纹现象	精度高,高分辨率,可动态测量
		刻度尺	有透射式和反射式	
	电压型	差动变压器	利用一次线圈、二次线圈及可动铁芯	可测 2~几十 mm,精度较高
	电磁感应型	直线感应同步器	利用电磁感应原理	具有高精度,高分辨率,抗干扰力强,使用寿命长,可用于长距离测量
	磁电型	磁尺	利用漏磁通变化产生感应电动势	可测位移达 3m
非直线型	非接触型	激光测距仪	利用激光原理	非接触测量,精度较高
		超声波测距仪	利用超声波原理	
		增量型编码器	利用脉冲发生器加计数器	
	光电型	绝对型编码器	输出信号是数字代码	精度高,使用方便
		圆光栅	利用光栅的莫尔条纹现象	精度高,高分辨率,可动态测量
		旋转变压器	利用电磁感应原理	
	电磁感应型	感应同步器	利用电磁感应原理	
		磁电型	磁尺	利用漏磁通变化产生感应电动势
	电阻型	电位计	利用电阻变化测量角度	结构简单,使用方便

5.3 速度、加速度传感器

5.4 力、力矩传感器和其他传感器

常用的速度、加速度传感器如表 47.6-5 所示。

表 47.6-6 列出了力、力矩及其他物理量传感器。

表 47.6-5 常用的速度、加速度传感器

名称	工作原理	特点
测速发动机	有交流和直流两种。用发电机原理	方便,有较高检测灵敏度
差动变压器式速度传感器	利用磁芯移动产生感应电势	
光电式速度和转速传感器	被测物移过光电池而输出阶跃电压信号	方便
霍尔式转速传感器	利用霍尔效应,测得单位时间内脉冲数	
加速度传感器	利用惯性力产生位移的电感或电容变化	

表 47.6-6 力、力矩及其他物理量传感器

传感器名称	工作原理	特点
力、力矩传感器	1. 利用应变片和应变杆测力、力矩 2. 利用磁分度圆和应变轴测力矩 3. 压电晶体的电荷变化与作用力成正比	测量范围较大 测量精度较高 测量范围宽,线性好,可动态测量
温度传感器	1. 基于热电效应所产生的热电势 2. 根据导体或半导体的电阻随温度而变化 3. 光学高温计单色辐射强度随温度升高而增加的原理	被测温度可较高,精度高 精度高,范围广 用于非接触式

5.5 传感器与微机的接口

量。与此相应的有三种基本接口方式,见图 47.6-7。

输入到微机的信息必须是微机能够处理的数字量。根据模拟量转换输入的精度、速度与通道等因素信息。传感器的输出形式可以分为模拟量、数字量和开关量。可用四种转换输入方式,如表 47.6-8 所示。

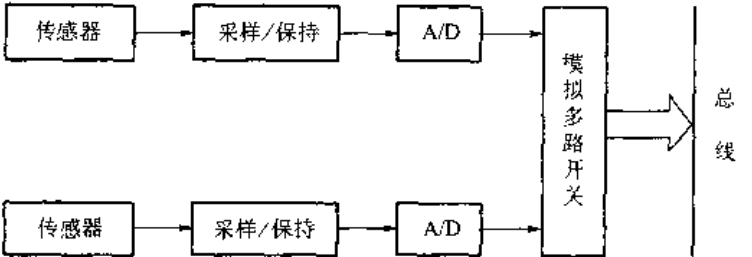
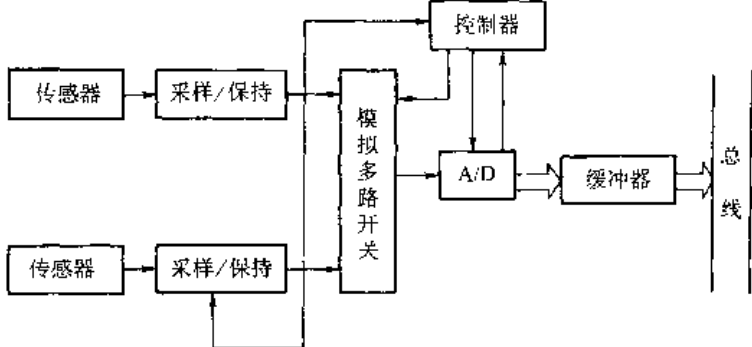
表 47.6-7 传感器与微机的基本接口

接口方式	基本方法
模拟量接口方式	传感器输出信号→放大→采样保持→模拟多路开关→A/D 转换 I/P 接口→微机
开关量接口方式	开关型传感器输出二值信号(逻辑 1 或 0)→三态缓冲器→微机
数字量接口方式	数字型传感器输出数字量(二进制代码、脉冲序列等)→计数器→三态缓冲器→微机

表 47.6-8 模拟量转换输入方式

类型	组成原理框图	特点
单通道直接型	<pre>                     graph LR                         Sensor[传感器] --&gt; AD[A/D]                         AD --&gt; Buffer[三态缓冲器]                         Buffer --&gt; Out1(( ))                         Buffer --&gt; Out2(( ))                     </pre>	形式简单,受转换电压幅度与速度的限制,应用范围窄

(续)

类型	组成原理框图	特点
多通道一般型		<p>节省元、部件, 速度低, 不能获得同一瞬时的各通道的模拟信号</p>
多通道同步型		<p>各采样/保持保持同时动作, 可测得在同一瞬时各传感器输出的模拟信号</p>
多通道并行输入型		<p>各通道直接进行转换, 送入微机或信号通道。灵活性大, 抗干扰能力强。根据传感器输出信号的通道可采用采样/保持或不同精度 A/D</p>

### 5.6 检测传感系统的设计原则

检测传感系统在机电一体化系统中起着十分重要的作用, 它将机电一体化系统自身的、作业对象的以及作业环境的工作状态参数进行检测, 并将检测到的状态参数信息提供给信息处理及控制系统进行信息处理并发出控制信号以控制广义执行机构。因此, 检测传感系统的设计原则:

- 1) 决定需要检测的工作状态参数;
- 2) 选择合适的检测传感器类型;
- 3) 确定检测传感器的安装位置及信号引出方式;
- 4) 设计传感器与微机接口方式和检测传感器模拟量转换输入方式。

## 6 信息处理及控制子系统的设计

### 6.1 信息处理及控制子系统的基本构成

信息处理及控制子系统是机电一体化系统中采用微机的控制器。因此, 信息处理及控制子系统的基本构成

可表示为图 47.6-22 所示虚线左边所示。

作为一种微机控制系统, 它的基本构成与微机类型、接口形式、控制方案和控制算法有关。

在确定信息处理及控制子系统的构成时应考虑如下两个问题:

#### (1) 控制系统的专用性与通用性的选择

对于大批量生产的机电一体化产品, 其控制系统采用专用控制系统, 专用控制系统可选用通用的 IC 芯片, 使其与广义执行机构和检测传感器相匹配。对于多品种、中小批量生产的机电一体化产品常常采用通用控制系统, 它选用微机并通过接口设计和软件编制来控制使控制系统达到专用化。

#### (2) 硬件与软件的权衡

对于运算与判断处理等功能适宜用软件来实现; 而其余不少的功能既可用硬件来实现, 又可用软件来实现。为了合理组成控制系统的硬件和软件, 通常根据系统的经济性和可靠性综合最优来确定。

在必须用分主元件组成硬件的情况下, 不如采用软件。如果能用通用的 LSI 芯片来组成所需的电路, 则

最好用硬件。

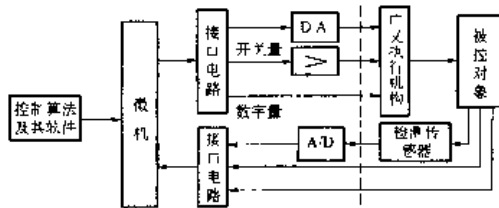


图 47.6-22 线性处理及控制子系统的基本构成

## 6.2 信息处理及控制子系统的一般设计过程及内容

不同的机电一体化系统对控制有不同的要求,控制子系统是围绕广义执行机构子系统的机械主功能需要进行设计的。因此,控制子系统的设计方法和步骤应按实际需要而定。但是其一般设计过程如图 47.6-23 所示。

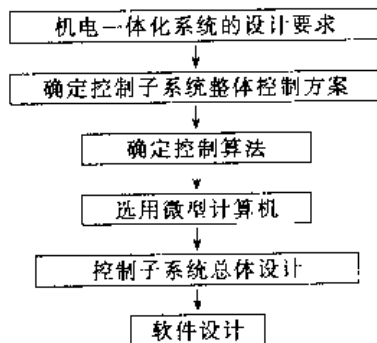


图 47.6-23 信息处理及控制子系统的一般设计过程

对于信息处理及控制子系统设计的主要内容有:

### (1) 确定控制子系统的整体方案

构思控制子系统的整体方案必须深入了解被控对象的控制要求,具体有:

1) 采用何种控制方式:采用开环控制结构比较简单但精度不高;采用闭环控制结构比较复杂,但精度较高。采用闭环控制应考虑采用何种检测传感器、检测精度要求如何。

2) 应考虑驱动元件的类型和执行机构的类型。

3) 应考虑对可靠性、精度和快速性有什么要求。

4) 考虑微机在整个控制系统中的作用,是设定计算、直接控制还是数据处理,微机应承担哪些任务,为完成这些任务,微机应具备哪些功能,需要哪些输入/输出通道,配备哪些外围设备。

5) 画出控制系统组成的初步框图,作为下一步设计的依据。

### (2) 确定控制算法

应对控制系统建立数学模型,确定其控制算法。所谓数学模型就是系统动态特性的数学表达式。它反映了系统输入、内部状态和输出之间的数量和逻辑关系。这些关系式为计算机进行运算处理提供了依据,即由

数学模型推出控制算法。所谓计算机控制,就是按照规定控制算法进行控制。因此,控制算法决定了控制系统的优劣。应根据不同的控制对象、不同的控制指标要求选择不同的控制算法。对于复杂的控制系统,其算法也较复杂,使控制较难实现。为此进行某些合理简化,忽略某些影响因素,使控制算法简化又能获得较好的控制效果。

### (3) 选择微型计算机

从控制的角度出发,微型计算机应能满足具有较完善的中断系统,具有实时控制性能,保证控制系统满足生产中提出的各种控制要求。

要有微型计算机并应有足够的存储容量,包括内存容量和外存储器,以保证存放程序和数据。

要有完备的输入/输出通道和实时时钟,以实现多种信息交换和按规定时间程序完成各种操作。

微型计算机包括单片机、单板机和 PC 计算机。

### (4) 控制系统总体设计

控制系统要综合考虑硬件和软件措施,解决微型机、被控对象和操作人员三者的信息交换的通路和分时控制的时序安排问题。保证系统能正常地运行。通过总体设计画出系统的具体构成框图。

### (5) 软件设计

微机控制系统的软件主要分为系统软件和应用软件。系统软件包括操作系统、诊断系统、开发系统和信息处理系统,通常这些软件不需用户设计,对用户来说,基本上只需了解其大致原理和使用方法就行了。而应用软件都要由用户自行编程,所以软件设计主要是应用软件的设计。

控制系统对应用软件的要求是具有实时性、针对性、灵活性和通用性。

应用软件的设计方法有两种,即模块化程序和结构化程序。模块化是将整个程序分成若干模块,一个模块完成一定的功能。结构化就是限定采用规定的结构类型和操作顺序,便于查找错误和纠正错误。

## 6.3 微型计算机的种类

微型计算机是微处理机(微处理器)、微型计算机、微型计算机系统的统称。它可按组装形式、微处理机位数以及用途来分类。

### (1) 按组装形式分类

1) 单片机 在一块集成电路芯片(LSI)上装有 CPU、ROM、RAM 以及输入/输出端口电路,该芯片称为单片微型计算机,简称单片机。

2) 单板机 将微型计算机的基本体系 CPU、一定容量的 ROM 和 RAM、输入/输出端口(I/O)以及一些辅助电路分别作成 LSI 芯片,并将它们配置在一块印

刷电路板上,用电缆线和外部设备直接联接起来,称为单板微型计算机,简称单板机。

3) 微型计算机系统 将微型计算机、ROM、RAM、I/O 接口电路、电源等组装在不同的印刷电路板上,然后组装在一个机箱内,再配上键盘、CRT 显示器、打印机、硬盘、软盘驱动器等多种外围设备和足够的系统软件,就构成了一个完整的微机系统。

(2) 按微处理机位数分类

按微处理机位数可将微型计算机分为 4 位、8 位、16 位、32 位和 64 位等机种。所谓位数是指数处理机并行处理的数据位数,即可同时传递数据的总宽度。

4 位机目前多做单片机,主要用作单机控制。8 位机有单片和多片之分,主要用于控制和计算。16 位机功能更强、性能更好,用于比较复杂的控制系统。

(3) 按用途分类

按用途可将微型计算机分为控制用和数据处理用微型计算机。对单片机来说,可分为通用型和专用型。

通用型单片机,即通常的各种系列的单片机。用户可按自己需要来设计接口和编制程序,可用作系统或产品的微控制器。

专用单片机也称专用微控制器,是针对某一特定产品或某一领域而开发的单片机,其内部系统结构或指令系统都是特殊设计(甚至内部已固化好程序)。

### 6.4 主要控制方法

在机电一体化系统中操作过程控制目的,一是根据操作条件的变化,制定最佳操作方案,二是对操作过程进行自动检测和自动控制,提高控制性能,实现规定的目的功能。

(1) 伺服控制系统

伺服控制系统是以机械位置或角度作为控制对象的自动控制系统。其输出量随输入量的变化而变化。

图 47.6-24 所示为位置控制伺服系统实例。被控对象为机器人的手臂,被控制量是手臂的转角。手臂转角的目标值  $\theta_1$  由指令电位器转换为电压  $u_1$ ,手臂的实际输出转角  $\theta_2$  由检测电位器检测,并转换为电压  $u_2$ ,只要电压  $u_2 \neq u_1$ ,即  $\theta_2 \neq \theta_1$  就有偏差电压  $e$  存在,该偏差电压经放大器放大后,驱动元件——伺服电机转动并通过齿轮减速机构带动机器人手臂转动,从而使输出转角  $\theta_2$  随  $\theta_1$  变化。该位置控制系统的工作过程可用图 47.6-25 所示框图表示。

在图 47.6-24 所示的系统,其输出量不仅受输入量控制,而且反馈回来影响输入量,所以称之为闭环控制或反馈控制。

如果在图 47.6-24 中不存在反馈电压  $u_2$ ,输出量与输入量之间只有顺向作用,而无反向联系的控制,此时的控制称开环控制。

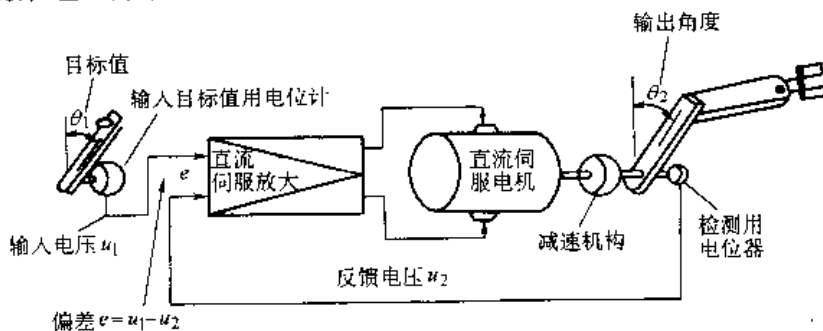


图 47.6-24 位置的电气伺服系统

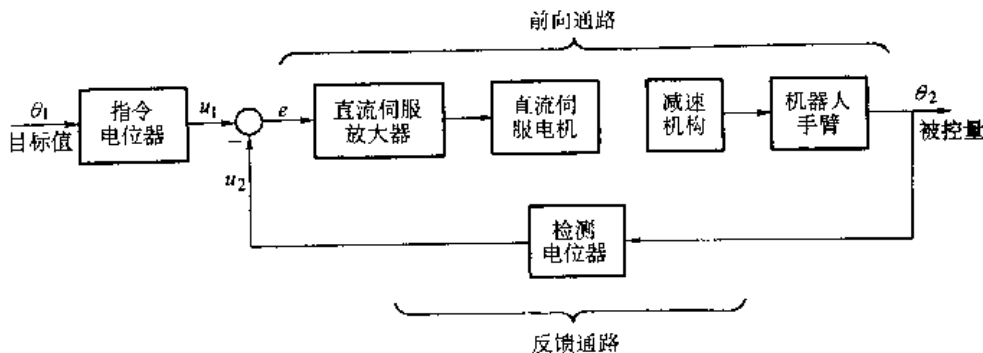


图 47.6-25 反馈控制系统框图

(2) 采样控制

在系统中只要有 一处的信号是脉冲信号或数字信号时,此系统称离散系统或采样系统。

所谓采样是指将连续时间信号转变为脉冲或数字信号的过程。

图 47.6-26 所示为炉温控制系统原理图。当炉温  $\theta$

偏离给定值时,测温电阻的阻值变化,使电桥失去平衡,检流计指针发生偏转,偏角为  $s$ 。检流计是一个高灵敏度的元件。同步电动机通过减速器带动凸轮,使检流计指针周期性上下运动,指针每隔  $T$  秒与电位器接触一次,每次接触时间为  $\tau$  秒。 $T$  称为采样周期, $\tau$  称为采样持续时间。当炉温连续变化时,电位器的输出是一串宽度为  $\tau$  的脉冲电压信号  $e_s^*(t)$ ,见图 47.6-27a。 $e_s^*(t)$  经放大器、电动机、减速器去控制阀门角度  $\varphi$ ,改变加热气体的进气量,使炉温趋向于给定值。

在大多数实际应用中,采样保持时间  $\tau$  远小于采样周期  $T$ ,可认为  $\tau$  趋近于零,即把采样器的输出近似地看成一串强度等于矩形脉冲面积的理想脉冲  $e_h^*(t)$ ,如图 47.6-27b 所示。

数字控制系统是指在系统中含有数字计算机或数字式控制器的系统。由于数字控制系统也需要采样,所以也称为采样控制系统。图 47.6-28 所示为数字控制

系统工作原理。其中对信号的采样和实现式分别由模数转换(A/D)装置和数模转换(D/A)装置实现的。

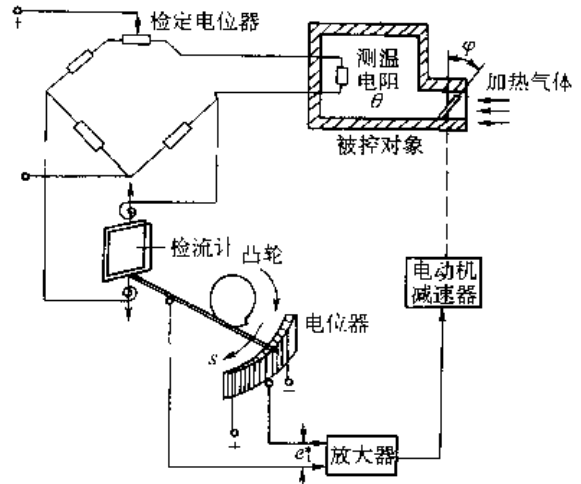


图 47.6-26 炉温控制系统原理图

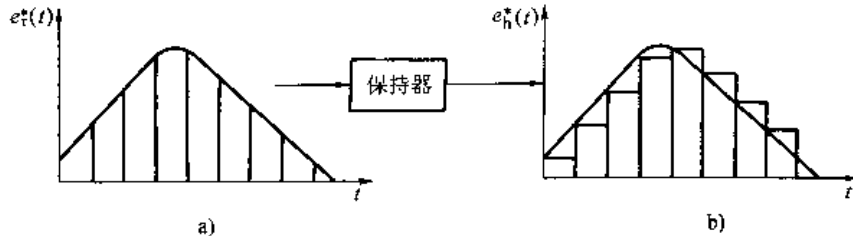


图 47.6-27 电位器的输出电压

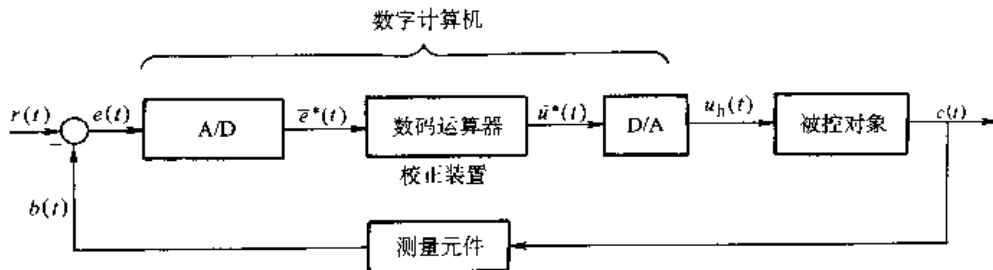


图 47.6-28 数字控制系统工作原理框图

### 6.5 控制系统的种类

#### (1) “量”控制与“逻辑”控制

“量”控制是以物理量的大小为控制对象的控制。“逻辑”控制是以物体“有”、“无”、“动”、“停”等逻辑状态为控制对象的控制。

“量”控制又分为“模拟控制”和“数字控制”。前者将物理量变换成大小与之对应的电压或电流等模拟量进行信号处理的控制。后者是将处理的“量”变换成数字量进行信号处理的控制。

“逻辑”又称顺序控制。

#### (2) 开环控制与闭环控制

在以数量大小、精度高低为对象的控制系统中,经常检测控制输出的结果与指令输入(目标值)进行比较,如有误差则进行自动修正,这种控制方式称为闭环

控制。与此不同,虽给出目标值输入,但不管输出结果如何的控制方式称为开环控制。

闭环控制又称反馈控制,它具有改善自动修改误差的稳定性和改善瞬态响应特性。

#### (3) 连续控制与非连续控制

在控制系统中,输入与输出在时间上保持连续关系的控制称连续控制。而在控制系统中输入与输出在时间上具有不连续关系的控制称非连续控制。

#### (4) 线性控制与非线性控制

按控制系统中是否存在非线性元件而分为线性控制或非线性控制。利用微机控制很易实现所需的非线性特性,如图 47.6-29 所示,这是在位置反馈控制系统中其目标位置阶跃变化时的响应特性。此时若增大反馈控制系统的放大率(增益),响应虽加快,但超调量变大会出现不稳定;反之,虽然超调量变小而比较稳定,



但响应变慢,这就是线性控制系统的缺点。利用微机进行非线性控制,随着接近目标值而逐渐提高放大率,其响应特性曲线比较理想。

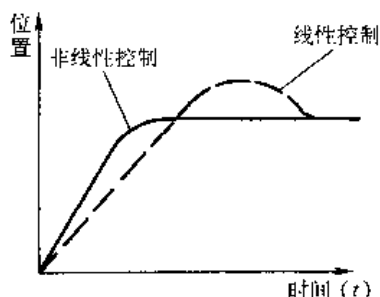


图 47.6-29 位置反馈控制系统的响应特性

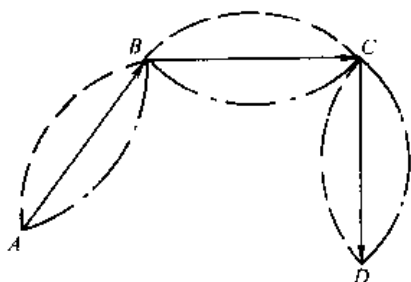


图 47.6-30 点位控制

(5) 点位控制和轨迹控制

点位控制(P. T. P——point to point)是只求起点和终点坐标的准确性,而不管经过任何路径的控制方式,如图 47.6-30 所示。

轨迹控制(C. P. ——continuous path)是需要对经过路径(轨迹)进行控制的控制方式,如图 47.6-31 所示。

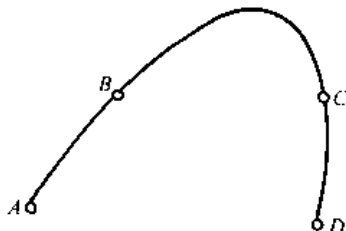


图 47.6-31 轨迹控制

6.6 动作控制方式及其特点

机电一体化系统中,动作控制方式是指执行机构从一点移动到另一点的过程中,对位置、速度、加速度等的控制方式。

(1) 位置控制方式

位置控制方式按其控制指令可分为绝对值方式和增量方式两种。绝对值方式是先确定基本坐标系,以此坐标系的坐标值为位置控制指令。而增量控制方式则以当前位置向下一个位置移动所需的移动量为控制指令。

1) 利用步进电机定位 以步进电机为驱动元件,把对应于所需移动量的脉冲数驱动步进电机进行定位,常用于定位精度要求不太高的场合。

2) 利用直流(或交流)伺服电机定位(绝对值方式) 图 47.6-32 所示为绝对值直流伺服电机定位的结构组成。采用反馈控制可以获得高速、高精度的定位。

3) 利用直流(或交流)伺服电机定位(增量式) 图 47.6-33 所示为增量式直流伺服电机定位的结构组成图。

(2) 速度控制方式

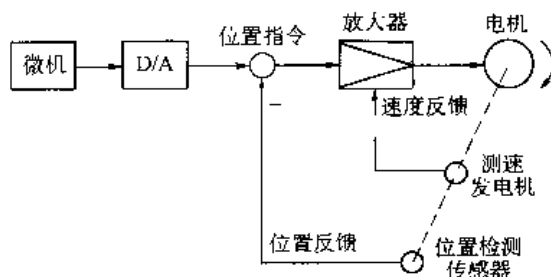


图 47.6-32 绝对值直流伺服电机定位

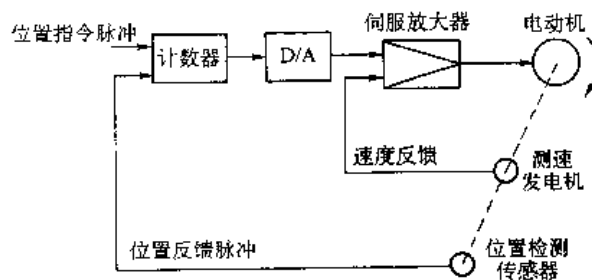


图 47.6-33 增量式直流伺服电机定位

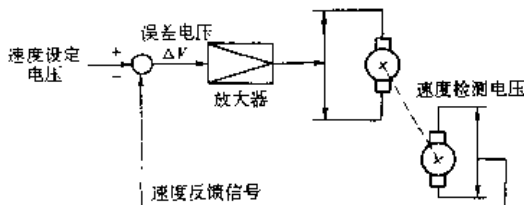


图 47.6-34 直流电机速度的模拟反馈控制

1) 速度的模拟反馈控制 图 47.6-34 所示为速度的模拟反馈控制的结构组成。其工作原理是利用电压比较电路,用设定电压与测速发电机的输出电压之差来求出设定转速与实际转速之差进行速度控制。

2) 速度的数字反馈控制 图 47.6-35 所示为速度的数字反馈控制的结构组成。其伺服放大器的输出和输入脉冲与速度反馈脉冲的相位差  $\alpha$  成正比,通过控制使其相位达到一致,从而达到速度控制的目的。

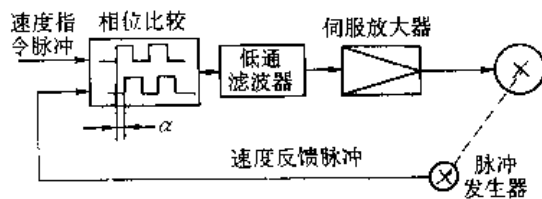


图 47.6-35 速度的数字反馈控制

## 7 机电一体化系统设计示例

机电一体化系统设计相对于纯粹机械系统设计要复杂得多,但是可以按机械主功能的需要,以各个执行动作为核心按 3 个子系统进行集成化设计。现用两个设计示例加以说明,希望读者能举一反三,灵活运用。

### 7.1 线料自动切断机

线料自动切断机是一种将连续线料切断成所需长度的机器。线料通过滚轮进行导引输送。根据总功能要求,线料自动切断机可以有两个执行动作——电缆间歇送料和电缆周期性切削。因此,可由间歇送料机构和电缆定时切削机构组成此自动切断机。由于定时切削机构需有较大切断力,一般可采用常速电机通过减速带动的曲柄滑块机构。而间歇送料机构由于需考虑切断长度多变,而且送料力又不大,因此采用机电一体化系统是十分合适的,我们对此作详细讨论。

#### (1) 广义执行机构的设计

1) 执行动作的要求 间歇送料机构的执行动作应是间歇移动,其移动位移是可变的。因此,采用机电一体化执行机构是较易实现这种要求的。

2) 广义执行机构的类型选择 广义执行机构是驱动元件与机构的集成体。它的运动转换应有移动→间歇移动和转动→间歇移动。具体来说可由图 47.6-36 来表示其组成方式。根据组成原理,它的方案至少有 9 种广义执行机构方案。通过对直线电机、伺服电机、步进电机的控制得到行程变化的间歇移动。

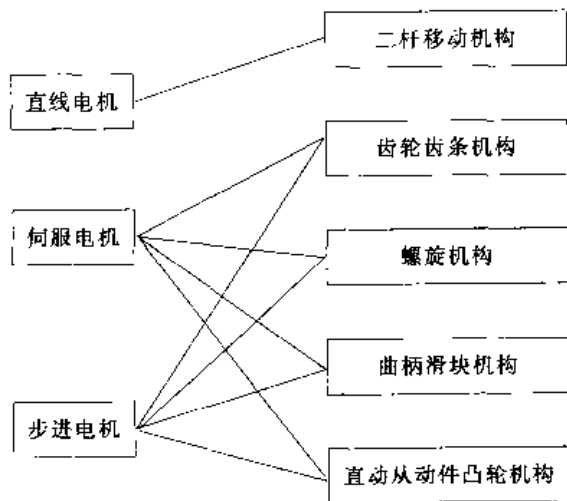


图 47.6-36 广义执行机构组成方案

#### (2) 传感器、控制系统与广义执行机构的集成

以步进电机-齿轮齿条组成的广义执行机构为例,其传感器、控制系统与广义执行机构的集成方案如图 47.6-37 所示。

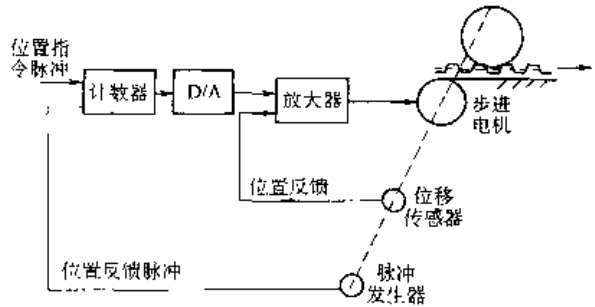


图 47.6-37 传感器、控制系统与广义执行机构的集成方案

它的传感器采用位移传感器,控制系统采用增量式脉冲反馈控制,最后使齿条实现间歇变行程的移动。

#### (3) 进行集成方案的评价决策

集成方案的评价决策是十分重要的。一般先可按工作精度、运动学和动力学要求、价格等方面对各子系统进行评价,然后对集成方案进行综合评价,最后进行决策。

(4) 对机电一体化系统进行动态分析 选对机电一体化系统建立数学模型,然后计算动态响应特性曲线,评定系统的动态性能。

### 7.2 多功能缝纫机的横针机构

家用缝纫机的发展已有 150 年的历史,为了增加缝纫机的功能,近几十年来,家用多功能缝纫机已有逐步替代普通家用缝纫机的趋势。家用多功能缝纫机与普通家用缝纫机不同之处就是增加了一个横针机构。通过横针机构使机针左右摆动,按不同的摆动运动规律缝制出各种较为复杂的线迹来,使缝制品增加花式提高附加值。

老的横针机构是采用纯机械式,它是通过一定速比的机械传动和花模凸轮来实现机针的摆动运动。为了改变线迹花样必须改变凸轮廓线形状,因此就形成了一个线迹花样更换一个花模凸轮。机械式横针机构具有使用不便、制造困难、输出柔性较差等缺点。因此,目前采用机电一体化系统作为横针机构,这种横针机构具有输出柔性好,操作方便等优点。

#### (1) 广义执行机构的设计

1) 执行动作的要求 横针机构的结构组成如图 47.6-38 所示。它是将原来刺料机构的针杆滑道由固定的变成左右摆动。这种摆动要求一是摆幅可大可小;二是摆动规律可以任意设定。

2) 广义执行机构的类型选择 它的驱动元件和执行机构都是有多种型式,根据组合原理可以有很多类型的广义执行机构,如图 47.6-39 所示。从原则上来说,它有 10 种类型以上,可以实现上述执行动作的要求。

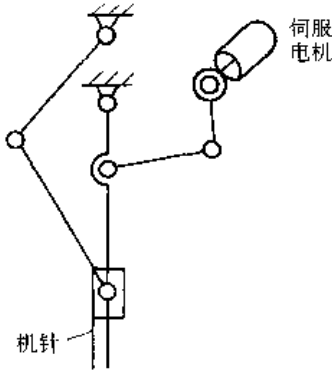


图 47.6-38 横针机构

(2) 传感器、控制系统与广义执行机构的集成

以伺服电机—齿轮机构组成的广义机构为例。

其控制系统与广义执行机构的集成方案如图 47.6-40 所示。通过微机发出指令脉冲使输出齿轮的转角运动规律符合规定要求。

(3) 对集成方案进行评价决策

集成方案数目可以很多,通过给定的评价指标体系进行综合评价,最后确定综合最优的集成方案。

(4) 对机电一体化系统进行动态分析

先建立机电一体化系统的数学模型,然后计算相应的动态特性曲线,评定系统的动态性能的优劣。

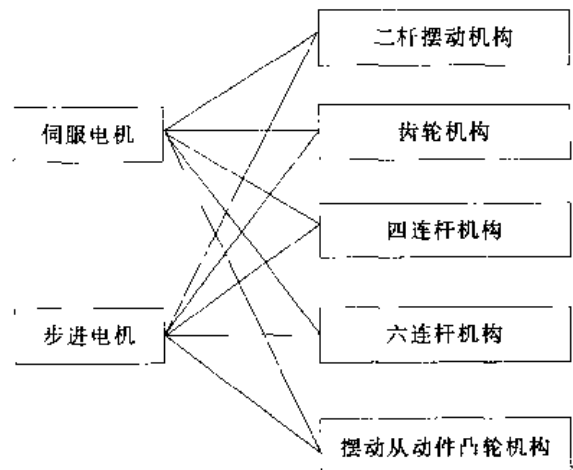


图 47.6-39 广义执行机构的组成方案

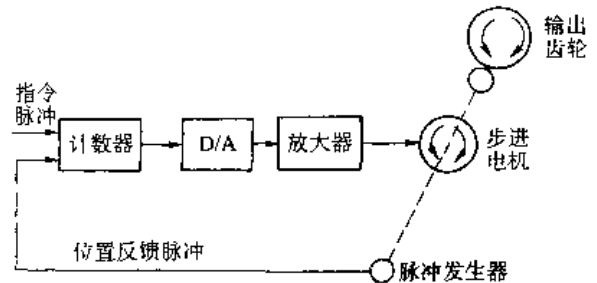


图 47.6-40 控制系统与广义执行机构的集成方案

# 第7章 机械运动系统的评价体系和评价方法

## 1 评价指标体系的确定原则

机械运动系统方案的构思和拟定的最终目标是最优地确定某一机械运动系统方案,并进一步解决机构系统设计问题。如何通过科学的评价和决策方法来确定最佳机械运动系统方案是机械运动方案设计的一个重要阶段。为此,必须根据机械运动方案的特点来确定评价特点、评价准则和评价方法,从而使评价结果更为准确、客观、有效,并能为广大工程技术人员认可和接受。

机械运动系统方案设计是机械设计的初始阶段的设计工作,其评价具有如下特点:

1) 评价准则应包括技术、经济、安全可靠三个方面的内容。这一阶段的设计工作只是解决原理方案和机构系统的设计问题,不具体地涉及机械结构设计的细节。因此,对经济性评价往往只能从定性角度加以考虑。对于机械运动系统方案的评价准则所包括的评价指标总数不宜过多。

2) 在机械运动系统方案设计阶段,各方面的信息一般来说都还不够充分,因此一般不考虑重要程度的加权系数。但是,为了使评价指标有广泛的适用范围,对某些评价指标可以按不同应用场合列出加权系数。例如承载能力,对于重载的机器应加上较大的加权系数。

3) 考虑到实际的可能性,一般可以采用0~4的五级评分方法来进行评价,即将各评价指标的评价等级分为五级。

4) 对于相对评价价值低于0.6的方案,一般认为较差,应该予以剔除。若方案的相对评价价值高于0.8,那么,只要它的各项评价指标都较均衡,则可以采用。对于相对评价价值介于0.6~0.8之间的方案,则要进行具体分析,有的方案在找出薄弱环节后加以改进,可成为较好的方案而被采纳。例如,当传递相对较远的两平行轴之间的运动时,采用V带传动是比较理想的方案。但是,当整个系统要求传动比十分精确,而其他部分都

已考虑到这一点而采取相应措施时(如高精度齿轮传动、无侧隙双导程蜗杆传动等),V带传动就是一个薄弱环节。如果改成同步带传动后,就能达到扬长避短的目的,又能成为优先选用的好方案。至于有的方案,确实缺点较多,又难以改进,则应予以淘汰。

5) 在评价机械运动系统方案时,应充分集中机械设计专家的知识 and 经验,特别是所要设计的这一类机器的设计专家的知识 and 经验;要尽可能多地掌握各种技术信息和技术情报;要尽量采用功能成本(包括生产成本和使用成本)指标值进行机械运动方案的比较。通过这些措施才能使机械运动方案评价更加有效。

因此,为了使机械运动系统方案的评价结果尽量准确、有效,必须建立一个评价指标体系,它是一个机械运动方案所要达到的目标群。对于机械运动方案的评价指标体系,一般应满足以下基本要求:

1) 评价指标体系应尽可能全面,但又必须抓住重点。它不仅要考虑到对机械产品性能有决定性影响的主要设计要求,而且应考虑到对设计结果有影响的主要条件。

2) 评价指标应具有独立性,各项评价指标相互间应该无关。这也就是说,采用提高方案中某一评价指标评价值的某种措施,不应对其他评价指标的评价值有明显影响。

3) 评价指标都应进行定量化。对于难以量化的评价指标可以通过分级量化。评价指标定量化有利于对方案进行评价与选优。

## 2 评价指标体系

### 2.1 机构的评价指标

机械运动系统方案是由若干个执行机构来组成的。在方案设计阶段,对于单一机构的选型或整个机构系统(机械运动系统)的选择都应建立合理、有效的评价指标。从机构和机构系统的选择和评定的要求来看,主要应满足5个方面的性能指标,具体见表47.7-1。

表 47.7-1 机构系统的评价指标

序号	1	2	3	4	5
性能指标	机构功能	机构的工作性能	机构的动力性能	经济性	结构紧凑
具体内容	1) 运动规律的形式 2) 传动精度	1) 应用范围 2) 可调性 3) 运转速度 4) 承载能力	1) 加速度峰值 2) 噪声 3) 耐磨性 4) 可靠性	1) 制造难易程度 2) 制造误差敏感度 3) 调整方便性 4) 能耗	1) 尺寸 2) 重量 3) 结构复杂性

确定这5个方面17项评价指标的依据,一是根据机构及机构系统设计的主要性能要求;二是根据机械设计专家的咨询意见。因此,随着科学技术的发展、生产实践经验的积累,这些评价指标需要不断增删和完善。有了比较合适的评价指标,将有利于评价选优。

### 2.2 几种典型机构的评价指标的初步评定

在构思和拟定机械运动方案时,相当多执行机构

往往首先选用连杆机构、凸轮机构、齿轮机构、组合机构这4种典型机构,这是因为这几种典型机构的结构特性、工作原理和设计方法都已为广大设计人员所熟悉,并且它们本身结构较简单,易于实际应用。表47.7-2对它们的性能和初步评价作简要评述,为评分和择优提供一定的依据。

如果在机械运动系统方案中采用自己创新的机构或其他的一些非典型机构,对评价指标应另作评定。

表 47.7-2 4种典型机构评价指标的初步评定

性能指标	具体项目	评 价			
		连杆机构	凸轮机构	齿轮机构	组合机构
功能 A	1. 运动规律形式	任意性较差,只能达到有限个精确位置	基本上能任意	一般作定速比转动或移动	基本上可以任意
	2. 传动精度	较高	较高	高	较高
工作性能 B	1. 应用范围	较广	较广	广	较广
	2. 可调性	较好	较差	较差	较好
	3. 运转速度	高	较高	很高	较高
	4. 承载能力	较大	较小	大	较大
动力性能 C	1. 加速度峰值	较大	较小	小	较小
	2. 噪声	较小	较大	小	较小
	3. 耐磨性	耐磨	差	较好	较好
	4. 可靠性	可靠	可靠	可靠	可靠
经济性 D	1. 制造难易程度	易	难	较难	较难
	2. 制造误差敏感	不敏感	敏感	敏感	敏感
	3. 调整方便性	方便	较麻烦	方便	方便
	4. 能耗	可靠	一般	一般	一般
结构紧凑 E	1. 尺寸	较大	较小	较小	较小
	2. 重量	较轻	较重	较重	较重
	3. 结构复杂性	简单	复杂	一般	复杂

(续)

### 2.3 机构选型的评价体系

机构选型的评价体系是由机械运动方案设计应满足的要求来确定的。依据上述评价指标所列项目,通过一定范围内的专家咨询,逐项评定分配分数值。这些分配分数值是按项目重要程度来分配的。这一工作是十分细致、复杂的。在实践中,还应该根据有关专家的咨询意见,对机械运动方案设计中的机构选型的评价体系不断进行修改、补充和完善。表47.7-3为初步建立的机构选型评价体系,它既有评价指标,又有各项分配分数值,正常情况下满分为100分。有了这样一个初步的评价体系,可以使机械运动系统方案设计逐步摆脱经验、类比的情况。

利用表47.7-3所示的机构选型评价体系,再加上对各个选用的机构评价指标的评价量化后,就可以对几种被选用的机构进行评估、选优。

表 47.7-3 初建的机构选型评价体系

性能指标	总分	项目	分配分	备 注
A	20	A1	15	以运动为主时,加权系数为1.5,即A×1.5
		A2	10	

性能指标	总分	项目	分配分	备 注
B	20	B1	5	受力较大时,在B3、B4上加权系数为1.5
		B2	5	
		B3	5	
		B4	5	
C	20	C1	5	加速较大时,加权系数为1.5,即C×1.5
		C2	5	
		C3	5	
		C4	5	
D	20	D1	5	
		D2	5	
		D3	5	
		D4	5	
E	15	E1	5	
		E2	5	
		E3	5	

### 2.4 机构评价指标的评价量化

利用机构选型评估体系对各种被选用机构进行评估、选优的重要步骤就是将各种常用的机构就各项评价指标进行评价量化。通常情况下各项评价指标较难量化,一般可以按“很好”、“好”、“较好”、“不太好”、“不好”五档来加以评价,这种评价当然应出自机械设计专

家的评估。在特殊情况下,也可以由若干个有一定设计经验的专家或设计人员来评估。上述五档评价可以量化为 4、3、2、1、0 的数值。由于多个专家评价总是有一定差别的,其评价指标的评价值取其平均值,因此不再为整数。如果数值 4、3、2、1、0 用相对值 1、0.75、0.5、0.25、0 表示,其评价值的平均值也就按实际情况而定。有了各机构实际的评价值,就不难进行机构选型。这种选型过程由于依靠了专家的知识经验和经验,因此可以避免个人决定的主观片面性。

## 2.5 机构系统选型的评估方法

在机械运动系统方案中,实际上是由若干个执行机构进行评估后将各机构评价值相加,取最大评价值的机构系统作为最佳机构运动方案。除此之外,也可以采用多种价值组合的规则来进行综合评估。

机械运动方案的选择本身是一个因素复杂、要求全面的难题,采用什么样的机构系统选型的评估计算方法值得认真去探索。上面采用评价指标体系及其量化评估的办法是进行机械运动方案选择的一大进步,只要不断完善评价指标体系,同时又注意收集机械设计专家的评价值的资料,吸收专家经验,并加以整理,那么,就能有效地提高设计水平。

## 3 价值工程方法

价值工程是以提高产品实用价值为目的,以功能分析为核心,以开发集体智力资源为基础,以科学分析方法为工具,用最低的成本去实现机械产品的必要功能。

价值工程中功能与成本的关系是:

$$V = \frac{F}{C} \quad (47.7-1)$$

式中  $V$ ——价值;

$F$ ——功能;

$C$ ——寿命周期成本。

机械运动系统方案的评价可以按它的各项功能求出综合功能评价,以便从多种方案中合理地选择最佳方案。即以功能为评价对象,以金额为评价尺度,找出某一功能最低成本。

下面先分别说明产品的功能、产品的寿命周期成本、产品的价值以及产品价值评定的思考等的含义。

### 3.1 产品的功能

价值工程的根本问题,是摆脱以事物(产品结构)为中心的研究,转向以功能为中心的研究。功能是机械产品设计的出发点和依据。用户所要求的是特定的功能而不是具体的产品结构本身,结构本身只不过是实

现特定功能的一种手段。例如,间歇运动机构的改进,如果单从机械结构出发来研究,最终仍离不开原来的框框,如从实现间歇运动这一功能出发就可采用步进电机。因此,功能定义可以帮助设计者打破老框框,创造新机构。

功能是指机械产品所具有的特定用途和使用价值。对于机械运动方案来说,特定用途就是指实现某一特定工艺动作过程,使用价值就是指机械实现了功能所体现的价值。对某一执行机构来说,特定用途就是指实现某一工艺动作,使用价值就是此动作所体现的效果。

### 3.2 产品的寿命周期成本

产品的寿命周期成本是指产品自研究、形成到退出使用所需的全部费用。产品的寿命周期成本是生产成本  $C_v$  与使用成本  $C_u$  之和,即

$$C = C_v + C_u \quad (47.7-2)$$

用户为获得机械产品而用的购置费,称为生产成本  $C_v$ ;而用户在使用机械产品过程中所支付的各种使用费用,称为使用成本  $C_u$ 。

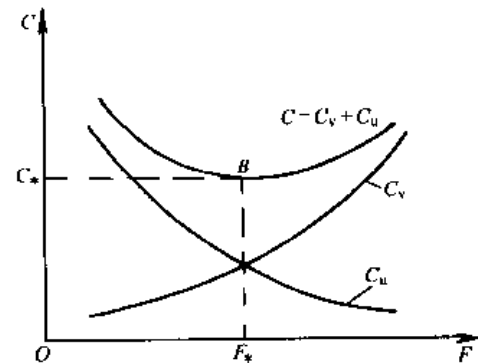


图 47.7-1 产品功能与产品成本间的关系

价值工程法的目的就是寻求不同的设计方案,以使最低寿命周期成本可靠地实现使用者所需功能,以获取最佳的综合效益。图 47.7-1 表示机械产品功能  $F$  与机械产品寿命周期成本  $C$  之间关系。在  $C$  曲线的最低点  $B$  处,产品寿命周期成本最低。价值工程追求的也是这一理想点。说明设计方案在技术、经济上更为合理。

### 3.3 产品的价值

为了评定机械产品的价值,必须使功能能够与成本进行比较。因此,功能也必须用货币来表示。每一机械产品都是为了实现用户需要的某种功能,为了获得这种功能必须克服某种困难,而克服困难的难易程度是可以设法用货币来表示的。这种用货币表示的实现功能的费用,亦即功能的货币表现,称为功能评价价值。

在大多数情况下,对于机械运动方案的功能有好几项,选择的分析对象为执行机构。例如家用缝纫机,它的四个执行机构——刺料机构、挑线机构、送料机构和勾线机构,它们的功能分别为 $F_1$ 、 $F_2$ 、 $F_3$ 、 $F_4$ 。

由此得出价值公式:

$$V = \frac{F_1 + F_2 + F_3 + F_4 + \dots}{C} \quad (47.7-3)$$

功能评价价值(即货币表示的功能)可以相加。

在评定机械运动方案的价值时, $V=1$ 表示实现功能所花的费用与其成本相适应,这是理想状况。 $V<1$ 表示实现功能的实际成本比其必需成本大,应该努力降低成本,使其趋近于1。 $V>1$ 表示用较少的成本实现了规定的功能,可以采取保持一定成本水平下适当地提高其功能。

### 3.4 机械运动方案的价值评定

价值工程评价过程,主要是功能成本分析和功能评价价值的确定。

#### (1) 功能成本分析

功能成本是实现功能所需费用,它包括生产成本和使用成本。对于机械运动方案的功能是由各执行机构来实现的,因此功能成本分析对象就是各个执行机构。功能成本分析主要依靠生产厂和用户的资料进行预测和估算,这就需要进行成本资料的积累和分析。这些工作往往有很强的针对性。例如,针对缝纫机、包装机械等进行资料积累。

#### (2) 功能评价价值的确定

从定义来看,功能评价价值是一个理论数值。在实际工作中,通常都是把功能目标成本作为功能评价价值。这一数值的确定,既要考虑用户的需求,还要考虑技术实现的可行性和经济性。确定这一数值的方法很多,下面介绍一种比较有效的方法——最低成本法。

最低成本法实际上是一种类比的方法,当功能的目标成本在理论上难以找到时,可以找出实际中实现同样功能的最低成本作为目标成本,具体做法为:

1) 广泛收集已有产品中完成同样功能的实际资料,弄清楚它们的功能相关条件,如工作性能、动力性能、经济性、结构紧凑等,并了解这些功能的实际满足程度和产品成本及用户反映等。

2) 统一产品的可比成本。将收集到的产品有关资料,按功能相关条件进行分类,功能及功能实现条件相似或相同的划分为一类,同一类中,依据功能满足程度再划分等级。

3) 根据成本资料估算出各自的功能成本。然后以产品功能实现程度为横坐标,以成本值为纵坐标绘制坐标图,把各产品实现该功能的情况画入坐标图,分别

描出“×”点。如图 47.7-2 所示。

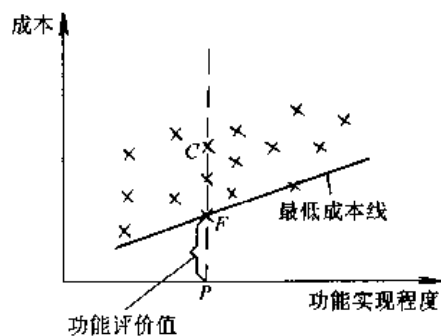


图 47.7-2 产品功能估算

4) 把图中最低点连成一条直线,这条直线就是按不同满足程度实现这一功能的最低成本线。从这条直线上可以很方便地求出目标成本。如图 47.7-2 中 P 点为原方案的满足程度, F 点代表满足同样功能水平的目标成本,而 C 点为目前成本。

这种方法要求有充分的实际数据作为依据,可靠性强、可比性好。而且由于目标成本在实际上是不断变化的,需要不断收集资料进行分析,适当地调整收集到的成本值。有了机械运动方案的功能成本和功能评价价值就可以进行几个机械运动方案的评估选优。但是,由于方案阶段不确定因素还比较多,因此困难较大。所以对某一种专门机械产品一定要在大量资料积累之后才能够有效地进行评价选择。此外,该方法由于强调机械的功能和成本,因此有可能对不同工作原理的方案进行评价,为人们进行方案创造开辟一条重要途径。

## 4 系统分析方法

系统分析法就是将整个机械运动系统方案作为一个系统,从整体上评价方案适合总的功能要求的情况,以便从多种方案中客观地、合理地选择最佳方案。系统工程评价是通过求总评价价值 $H$ 来进行的,通常 $Q$ 个方案中 $H$ 值最高的方案为整体最佳的方案。当然,最终决策还可以由设计者根据实际情况作出选择。例如,完成某一实际工艺动作有许多机械运动方案,有时为了

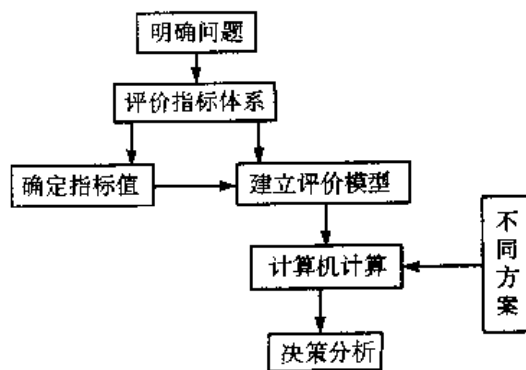


图 47.7-3 系统工程评价步骤

满足一些特殊的要求,并不一定要选择  $H$  值最高的方案,而是选择  $H$  值稍低而某些指标值较高的方案。

图 47.7-3 表示系统工程评价步骤的框图。

对于任一机械运动系统方案,要达到的目标很多,它们的要求也不一样,系统工程评价法就是将一个机构系统从整体上对其各项评价指标进行综合评价。

#### 4.1 系统工程评价方法的基本原则

为了达到机构系统从整体上进行综合评价,必须遵循以下几个原则:

##### (1) 要保证评价的客观性

系统综合评价的目的是为了决策和选优,因此评价的客观性、有效性和合理性必须充分保证。这就要求评价的依据要全面和可靠;评价专家要有一定的权威性和客观性;评价方法要合理和可靠等等。

##### (2) 要保证方案的可比性

各个供选择的机械运动方案在保证实现系统的基本功能上要有可比性和一致性。不能突出一点不及其余,要进行方案的全面比较,才能防止片面性和个人主观武断。

##### (3) 要有适合机械运动方案的评价指标体系

评价指标既要包括机构系统所要实现的定量目标,也要包括机构系统所应满足的定性要求。评价指标体系制定得好坏,对于评价结果的合理和有效性十分重要。评价指标体系的建立过程应充分集中领域专家的知识经验。

#### 4.2 建立评价指标体系和确定评价指标值

对于机械运动方案的评价指标体系如前所述,定为 5 个方面 17 项评价指标,从表 47.7-3 中看出这 17 项评价指标的重要程度按分配分的多少来定,如果在具体的机械运动方案中要考虑一些特殊情况,还可在有关项评价指标的分配分加权系数。

确定评价指标值的过程称之为量化,它是把具体某一执行机构所能达到评价指标要求的程度进行量化,一般采用相对比值办法,将实现程度定为 1、0.75、0.5、0.25、0。对完全能实现评价指标规定的要求的机构就定为 1,也就取得这项评价指标分配分的满分,否则就要将分配分打一个折扣。量化的方法通常有 3 种:直接量化法、间接量化法、分等级法。现在是采用了分等级法。

如何确定机构系统评价指标体系及其各项评价指标的分配分是机械运动系统方案评估中十分重要的步骤。这些工作要通过领域专家的咨询而最后确定下来。表 47.7-3 所示的就是——种集思广益的评价指标体系和各项分配分。

为了对各机械运动方案进行评估,还必须对各个具体的执行机构的各项指标的实现程度用相对比值来表示,这些相对比值的确定一定要根据机构的技术资料、手册、实验数据以及领域专家的知识经验来确定。如果由多名专家用填表方式来确定相对比值,其平均值就作为最后确定的相对比值。

#### 4.3 建立评价模型

评价模型应能综合考虑各评价指标,得出合理的评价结果。体现了系统工程评价法的具体计算原理。评价模型不但应考虑各指标在总体目标中的重要程度,还应考虑各指标之间的相互影响及结合状态。一般不能只用加权方法,还应运用多种价值组合规则。当各因素之间互相促进时用代换规则,当各因素之间可以互相补偿时用加法规则;当因素个个重要时用乘法规则。对于由 A、B、C 3 个执行机构组成的机械运动方案,如图 47.7-4 所示。

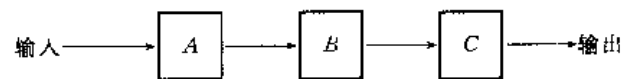


图 47.7-4 3 个执行机构组成的机械运动方案评价模型

它的总评价模型为  $H$ ;

$$H = (H_A^{\omega_A} \cdot H_B^{\omega_B} \cdot H_C^{\omega_C}) \quad (47.7-4)$$

式中  $\omega_A, \omega_B, \omega_C$  为加权因子,根据各执行机构 A、B、C 在整体中所占的重要程度而定。必须注意,运用乘法规则时的加权因子采用指数加权。

评价模型的结构如图 47.7-5 所示。

其中,

$H_A = (U_1(\cdot)U_2 \cdots (\cdot)U_N)$  为乘法规则;

$H_B = (U_{N+1}(+)U_{N+2} \cdots (+)U_P)$  为加法规则;

$H_C = (U_{P+1}(\cdot)U_{P+2} \cdots (\cdot)U_S)$  为乘法规则。

每个指标  $U_i$  又可由若干子指标组成,可根据设计要求采用某一运动规则来组成。对于加法规则:

$$U_i = \sum_{j=1}^M W_j \quad (47.7-5)$$

经过计算得出所有方案的评价值后,应对所得结果进行分析,选取其中最能适合设计要求的方案。例如, A 执行机构有  $m$  个方案、B 执行机构有  $n$  个方案、C 执行机构有  $p$  个方案,那么根据排列组合理论和实际可行性,此机械运动系统方案共有方案数为:

$$Q = mnp - k \quad (47.7-6)$$

式中  $k$  为 A、B、C 三个执行机构组成的不可行方案数。不可行方案主要是由于三个执行机构在五大类评价指标上不能匹配工作。



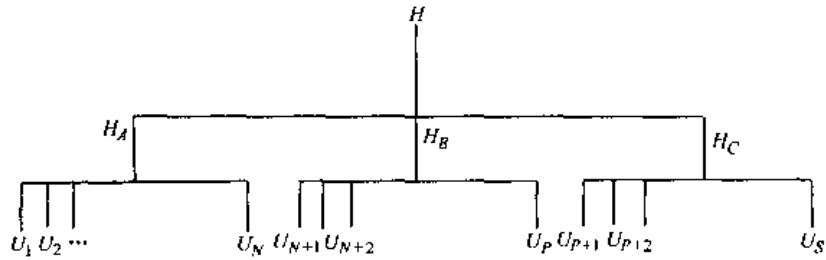


图 47.7-5 评价模型结构

在通常情况下,  $Q$  个方案中以  $H$  值最高的机械运动系统方案为整体最佳的方案。当然, 由系数工程方法算出的评价价值只是为设计者选择机械运动方案提供了可靠的依据。但是, 最终的选择方案的决策还可以由设计者根据实际情况作出最终选择。例如, 在实际工作中, 有时为了满足一些特殊的要求, 并不一定选择  $H$  值最高的方案, 而是选择  $H$  值稍低, 但某些指标值较高的方案。

### 5 模糊综合评价法

在机械运动系统方案评价时, 由于评价指标较多, 如有应用范围、可调性、承载能力、耐磨性、可靠性、制造难易、调整方便性、结构复杂性等等, 它们很难用定量分析来评价, 属于设计者的经验范畴, 只能用“很好”、“好”、“不太好”、“不好”等“模糊概念”来评价。因此, 应用模糊数学的方法进行综合评价将会取得更好的实际效果。模糊评价就是利用集合与模糊数学将模糊信息数值化, 以进行定量评价的方法。

#### 5.1 模糊综合评价中主要运算符号

- $\in$ : 表示元素与集合的属;  $\notin$  或  $\notin$ : 不属于;
- $\subseteq$ : 包含;  $\not\subseteq$ : 不包含;
- $\subset$ : 真包含;  $\cup$ : 并;
- $\cap$ : 交;  $\tilde{A}$ : 模糊集合;
- $\tilde{A}^c$ : 模糊集的补;  $\wedge$ : 取小运算;
- $\tilde{\vee}$ : 取大运算。

#### 5.2 模糊集合的概念

定义: 论域  $U$  中的模糊集合  $\tilde{A}$ , 是以隶属函数  $\mu_{\tilde{A}}$  为表征的集合, 即:

$$\mu_{\tilde{A}}: U \rightarrow [0, 1]$$

$$u \rightarrow \mu_{\tilde{A}}(u)$$

$\mu_{\tilde{A}}$  称作  $\tilde{A}$  的隶属函数,  $\mu_{\tilde{A}}(u)$  表示元素  $u \in U$  属于  $\tilde{A}$  的程度, 并称  $\mu_{\tilde{A}}(u)$  为  $u$  对于  $\tilde{A}$  的隶属度。

关于此定义, 有如下几点说明:

- 1)  $\tilde{A}$  的隶属函数与普通集合的特征函数相比, 它

是经典集合的一般化, 而经典集合则是它的特殊形式。亦即  $\tilde{A}$  是  $U$  上的一个模糊子集。

2) 模糊子集完全由其隶属函数来刻画。事实上, 我们可以建立模糊子集与隶属函数间的一一对应关系。

$\mu_{\tilde{A}}(u)$  接近于 1, 表示  $u$  隶属于  $\tilde{A}$  的程度大; 反之  $\mu_{\tilde{A}}(u)$  接近于零, 表示  $u$  隶属于  $\tilde{A}$  的程度小。

3) 隶属函数是模糊数学的最基本概念, 借助它我们才有可能对模糊集合进行量化, 也才有可能利用精确数学方法去分析和处理模糊信息。

隶属函数通常是根据经验或统计来确定, 它本质上是客观事物的属性, 但往往带有一定的主观性。正确地建立隶属函数, 是使模糊集合能够恰当地表现模糊概念的关键。所以, 应用模糊数学去解决实际问题, 往往归结为找出一个恰当的隶属函数, 这个问题解决了, 其他问题也就迎刃而解了。

为了说明隶属函数与其模糊集合的关系, 举例如下。

例: 设  $U = [0, 100]$  表示年龄的某个集合,  $\tilde{A}$  和  $\tilde{B}$  分别表示“年老”与“年轻”, 其隶属度函数分别见图 47.7-6 和图 47.7-7, 其表达式如下:

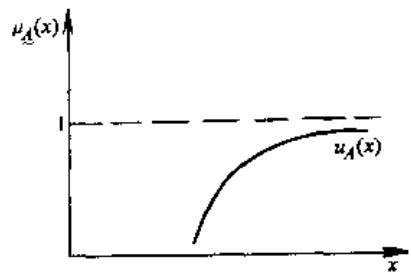


图 47.7-6 “年老”隶属度函数

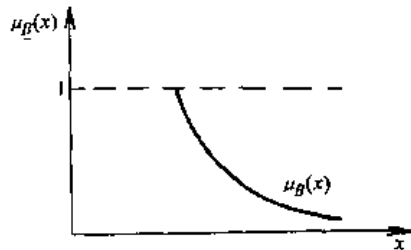


图 47.7-7 “年轻”隶属度函数

$$\mu_A(x) = \begin{cases} 0 & 0 \leq x \leq 50 \\ \left[ 1 + \left( \frac{x-50}{50} \right)^2 \right]^{-1} & 50 < x \leq 100 \end{cases}$$

$$\mu_B(x) = \begin{cases} 0 & 0 \leq x \leq 25 \\ \left[ 1 + \left( \frac{x-25}{25} \right)^2 \right]^{-1} & 25 < x \leq 100 \end{cases}$$

如果  $x=60$ , 则有  $\mu_A(60)=0.80, \mu_B(60)=0.02$ , 即是说 60 岁属于“年老”的程度为 0.80, 属于“年轻”的程度为 0.02, 故可以认为 60 岁是比较老的。

### 5.3 隶属度函数的确定方法

一个模糊集合在给定某种特性之后, 就必须建立反映这种特性所具有的程度函数即隶属度函数。它是模糊集合应用于实际问题的基石。一个具体的模糊对象, 首先应当确定其切合实际的隶属度函数, 才能应用模糊数学方法作具体的定量分析。

模糊评价的表达和衡量是用某一评价指标的评价概念(如优、良、差)隶属度的高低来表示。例如, 某方案的调整方便性, 一般不可能是绝对方便或绝对不方便, 而被认为对方便性的概念有八成符合, 那么就称它对调整方便性的隶属度为 0.8。

隶属度可采用统计法或通过已知隶属度函数求得。

#### (1) 模糊统计试验法

模糊统计试验, 是对评价指标体系中某一指标进行模糊统计试验, 其试验次数应足够多, 使统计得到的隶属频率稳定在某一数值范围, 由此求得较准确的隶属度。

例如, 为了对机械运动方案中某执行机构的调整方便性隶属度函数进行统计试验。由 20 位机械设计人员进行评定, 其数据见表 47.7-4。

由表 47.7-4 可见, 此指标在“好”处的隶属度为 0.75。

#### (2) 二元对比排序法

二元对比排序法确定隶属度, 在实际工作中, 常常能对不易量化的概念得到较好的数据处理, 但主观色彩较浓厚。下面介绍二元对比排序法中的择优比较法。它是经过抽样试验后, 利用统计方法求取隶属度的。

表 47.7-4 某机械执行机构调整方便性评价统计

序号	评价	频数	相对频数
1	很好	1	0.05
2	好	15	0.75
3	较好	3	0.15
4	不太好	1	0.05
5	不好	0	0

例如, 对于某种评价指标, 5 种机构哪种最好?

设论域  $U = \{\text{机构 I, 机构 II, 机构 III, 机构 IV, 机构 V}\}$ 。

从从事机械设计的科技人员中, 随机抽取 50 人, 每人被测 20 次, 每次在  $U$  中选两种机构对比, 被测者从两种机构中择优指定自己选定的机构。

每个被试者按表 47.7-5 中的次序反复进行两遍, 结果记于表 47.7-6 中。

择优比较法将表 47.7-6 各行数字相加, 按总和数值大小排序。百分数是由各行总和除以“ $\Sigma$ ”列总和后求得。其中各百分数就代表某评价指标“好”的隶属度。由表 47.7-6 可见, 机构 III 为最好。

表 47.7-5 择优选定记录

	机构 I	机构 II	机构 III	机构 IV	机构 V
机构 I					
机构 II	1				
机构 III	5	2			
机构 IV	8	6	3		
机构 V	10	9	7	4	

表 47.7-6 择优选定记录结果与排序

择优次数	I	II	III	IV	V	$\Sigma$	%	顺序
I		52	52	54	66	224	22.4	2
II	48		84	48	58	238	23.8	1
III	47	16		53	61	177	17.7	4
IV	45	52	47		64	208	20.8	3
V	40	52	39	22		153	15.3	5

### 5.4 模糊综合评价

机械运动方案的评价指标的评价往往是模糊的, 因此需采用模糊综合评价的方法对机构系统的方案作

出最佳决策。

#### (1) 确定评价因素集

评价因素集又称评价指标集, 其中每一因素都是评价的“着眼点”。

对于一个执行机构的评价因素集,由表 47.7-2 可得:

$$U = \{A, B, C, D, E\}$$

式中:  $A = (A_1, A_2); B = (B_1, B_2, B_3, B_4);$

$C = (C_1, C_2, C_3, C_4); D = (D_1, D_2, D_3, D_4);$

$E = (E_1, E_2, E_3)。$

为了全面评价某一选定的执行机构,它的评价指

标集应由专家群来确定,以力求全面、合理。

(2) 确定评价等级集合

对于 U 中的各因素作出评价等级,一般可以按“很好”、“好”、“较好”、“不太好”、“不好”五个等级来加以评价。因此,请 N 个专家,分别对 U 中各因素作出评价  $u_i$ ,列于表 47.7-7,其中评价因素集中的因素  $u_i$  有  $x_{ij}$  个专家评定为  $u_j$ 。

表 47.7-7 确定评价等级集合

	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$\Sigma$
$u_1(A_1)$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$N$
$u_2(A_2)$	$X_{21}$	$X_{22}$	$X_{23}$	$X_{24}$	$X_{25}$	$N$
$u_3(B_1)$	$X_{31}$	$X_{32}$	$X_{33}$	$X_{34}$	$X_{35}$	$N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$u_{17}(E_3)$	$X_{17-1}$	$X_{17-2}$	...	...	...	$N$

(3) 确定评价矩阵

对于某一执行机构都可确定从 U 到 V 的评价矩阵,亦可称为模糊关系 R:

$$\tilde{R} = (r_{ij})_{n \times m} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nm} \end{bmatrix}$$

式中  $r_{ij} = \frac{x_{ij}}{N}$ 。

对于一个执行机构,它的评价因素有 n 个,  $n=17$ ; 它的评价等级有 m 个,  $m=5$ 。

(4) 确定权数分配集

权数又称权重,它是表征各评价因素相对重要性大小的估测。权数分配集用 A 表示:

$$A = (a_1, a_2, a_3, \dots, a_n)$$

式中  $a_i > 0$ , 且  $\sum_{i=1}^n a_i = 1$ 。

权数确定方法很多,对于机械运动方案评估可以采用专家估测法。这种方法取决于机械设计领域中的专家的知识与经验,各评价指标的权数都可由专家群作出判断。

设评价指标集为  $U = \{u_1, u_2, u_3, \dots, u_n\}$ , 请 M 个专家分别就 U 中元素作出权数判定,其结果列于表 47.7-8。

表 47.7-8 专家对评价因素权数判定

专 家	评 价 指 标					$\Sigma$
	$u_1$	$u_2$	...	$u_n$		
	权 数					
专家 1	$a_{11}$	$a_{12}$	...	$a_{1n}$		1
专家 2	$a_{21}$	$a_{22}$	...	$a_{2n}$		1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
专家 M	$a_{M1}$	$a_{M2}$	...	$a_{Mn}$		1
$\frac{1}{M} \sum_{i=1}^M a_{ij} = t_i$	$\frac{1}{M} a_1$	$\frac{1}{M} a_2$	...	$\frac{1}{M} a_n$		1

显然,表中各行之和等于 1,即  $\sum_{j=1}^n a_{ij} = 1 (i = 1, 2, \dots, M)$ 。根据表 47.7-8,可取各评价因素权数的平均值作为其权数,表中  $a_i (i = 1, 2, \dots, n)$  表示  $\sum_{j=1}^M a_{ij}$ , 即各行之和,那么  $a_i$  对应于指标  $u_i$  的权数为:

$$t_i = \frac{1}{M} \sum_{j=1}^M a_{ij} = \frac{a_i}{M}$$

在实际确定权数过程中,为了使所得权数更加客观、合理,一般应剔除  $a_{kj} = M_{\max}(a_{ij})$  及  $a_{kj} = M_{\min}(a_{ij})$ ,

亦即除去一个最大值和一个最小值,然后将其余各值平均后得到权数  $t_i$ 。

由于表 47.7-3 中所列评价性能指标的分配是征集了专家意见后确定的,因此按分配分可得到各评价指标(评价因素)的权数,17 项评价指标的权数为:

$$\tilde{A} = (0.15, 0.10, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05)$$

(5) 计算模糊决策集

在确定评价矩阵 $R$ 和权数分配集 $A$ 以后,我们可以按下式求模糊决策集 $B$ 。

一般地可令

$$B = A \circ R \text{ (“} \circ \text{”为算子符号)}$$

$B$ 的算法主要有两种:

1)采用模糊矩阵的复合算法

$$B = A \cdot R = (b_1, b_2, b_3, \dots, b_m)$$

$$b_j = \bigvee_{i=1}^n (a_i \wedge r_{ij}) \quad (j=1, 2, \dots, m)$$

即“ $\circ$ ”取“ $\wedge$ ”、“ $\vee$ ”运算(即取小运算和取大运算)。

算)。

现以简单例子说明运算过程,设 $A = (0.25, 0.20, 0.20, 0.20, 0.15)$ ,方案评价矩阵 $R$ 可求得:

$$R = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.4 & 0.3 & 0.2 & 0 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.1 & 0.2 \end{bmatrix}$$

那么模糊决策集:

$$B = A \cdot R = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.4 & 0.3 & 0.2 & 0 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.1 & 0.2 \end{bmatrix}$$

$$\begin{aligned} &= ((0.25 \wedge 0.4) \vee (0.2 \wedge 0.4) \vee (0.2 \wedge 0.3) \vee (0.2 \wedge 0.3) \vee (0.15 \wedge 0.2), \\ & (0.25 \wedge 0.3) \vee (0.2 \wedge 0.3) \vee (0.2 \wedge 0.2) \vee (0.2 \wedge 0.3) \vee (0.15 \wedge 0.2), \\ & (0.25 \wedge 0.2) \vee (0.2 \wedge 0.2) \vee (0.2 \wedge 0.2) \vee (0.2 \wedge 0.1) \vee (0.15 \wedge 0.3), \\ & (0.2 \wedge 0.1) \vee (0.2 \wedge 0) \vee (0.2 \wedge 0.2) \vee (0.2 \wedge 0.2) \vee (0.15 \wedge 0.1), \\ & (0.25 \wedge 0) \vee (0.2 \wedge 0.1) \vee (0.2 \wedge 0.1) \vee (0.2 \wedge 0.1) \vee (0.15 \wedge 0.2)) \\ &= (0.25 \vee 0.2 \vee 0.2 \vee 0.2 \vee 0.15, 0.25 \vee 0.2 \vee 0.2 \vee 0.2 \vee 0.15, \\ & 0.2 \vee 0.2 \vee 0.2 \vee 0.1 \vee 0.15, 0.1 \vee 0 \vee 0.2 \vee 0.2 \vee 0.1, \\ & 0 \vee 0.1 \vee 0.1 \vee 0.1 \vee 0.15) \\ &= (0.25, 0.25, 0.2, 0.2, 0.15) \end{aligned}$$

评价结果表明,该方案“很好”的程度为 0.25,“好”的程度为 0.25,“较好”的程度为 0.2,“不太好”的程度为 0.2,“不好”的程度为 0.15。假如对 $B = (0.25, 0.25, 0.2, 0.2, 0.15)$ 进行归一化处理,即 $B^* = \left( \frac{0.25}{1.05}, \frac{0.25}{1.05}, \frac{0.2}{1.05}, \frac{0.2}{1.05}, \frac{0.15}{1.05} \right) = (0.233, 0.238, 0.190, 0.190, 0.144)$ ,就是说,认为该方案“很好”的占 23.8%，“好”的占 23.8%，“较好”的占 19%，“不太好”的占 19%，“不好”的占 14.4%。

这种方法因为采用了“ $\wedge$ ”、“ $\vee$ ”运算,对于某些问题,可能丢失了太多的信息,使结果显得粗糙。特别是评价因素较多,权数分配又较均衡时,由于 $\sum_{i=1}^n a_i = 1$ ,因而使每一个因素所分得的权重 $a_i$ 必然很小,于是利用“ $\wedge$ ”、“ $\vee$ ”运算时,使综合评价中得到的 $b_j$ 注定很

小( $b_j \leq \bigvee a_i$ )。这时较小的权数通过“ $\vee$ ”运算而被剔除了,那么实际得到的结果往往会掩盖所有评价因素的评价,而变得不够真实。因此需要采用以下的改进方法。

2)改进的运算方法

$$B = A \circ R = (b_1, b_2, b_3, \dots, b_m)$$

$$b_j = \sum_{i=1}^n (a_i \cdot r_{ij}) = (a_1 r_{1j}) \oplus (a_2 r_{2j}) \oplus \dots \oplus (a_n r_{nj}) \quad (j=1, 2, \dots, m)$$

即“ $\circ$ ”取“ $\bullet$ ”、“ $\oplus$ ”算子; $a \cdot b = a \cdot b$ 乘积算子;

$a \oplus b = (a+b) \wedge 1$  闭合加法算子。 $\sum_{i=1}^n$ 表示对几个数在下求 $\oplus$ 和。这种算法简记为 $M(\bullet, \oplus)$ 。

对上法中所列举例子用改进的运算方法来计算模糊决策集。

$$B = A \circ R = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.4 & 0.3 & 0.2 & 0 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.1 & 0.2 \end{bmatrix}$$

采用 $M(\bullet, \oplus)$ ,有

$$B = ((0.25 \times 0.4) \oplus (0.2 \times 0.4) \oplus (0.2 \times 0.3) \oplus (0.2 \times 0.3) \oplus (0.15 \times 0.2), \\ (0.25 \times 0.3) \oplus (0.2 \times 0.3) \oplus (0.2 \times 0.2) \oplus (0.2 \times 0.3) \oplus (0.15 \times 0.2),$$

$$\begin{aligned}
 & (0.25 \times 0.2) \oplus (0.2 \times 0.2) \oplus (0.2 \times 0.2) \oplus (0.2 \times 0.1) \oplus (0.15 \times 0.3), \\
 & (0.25 \times 0.1) \oplus (0.2 \times 0) \oplus (0.2 \times 0.2) \oplus (0.2 \times 0.2) \oplus (0.15 \times 0.1), \\
 & (0.25 \times 0) \oplus (0.2 \times 0.1) \oplus (0.2 \times 0.1) \oplus (0.2 \times 0.1) \oplus (0.15 \times 0.2) \\
 = & (0.1 \oplus 0.08 \oplus 0.06 \oplus 0.06 \oplus 0.03, 0.075 \oplus 0.06 \oplus 0.04 \oplus 0.06 \oplus 0.03, \\
 & 0.05 \oplus 0.04 \oplus 0.04 \oplus 0.02 \oplus 0.45, 0.025 \oplus 0 \oplus 0.04 \oplus 0.04 \oplus 0.015, \\
 & 0 \oplus 0.02 \oplus 0.02 \oplus 0.02 \oplus 0.03) \\
 = & (0.33, 0.265, 0.195, 0.12, 0.09)
 \end{aligned}$$

归一化处理后有

$$B^* = (0.33, 0.265, 0.195, 0.12, 0.09)$$

上述计算结果表明,认为方案“很好”的占33%，“好”的占26.5%，“较好”的占19.5%，“不太好”的占12%，“不好”的占9%。如果把认为方案叫“很好”、“好”和“较好”这三者相加就占了79%。

采用  $M(\wedge, \vee)$  与  $M(\bullet, \oplus)$  的计算结果不同,是运算算子不同的缘故。实际计算结果表明,当中元素较均衡时,利用  $M(\wedge, \vee)$  运算结果是失真的,但取  $M(\bullet, \oplus)$  则弥补了  $M(\wedge, \vee)$  算法的不足,所以实际工作中要根据不同情况注意选择运算算子。

### 6. 模糊综合评价

对于单一机构的选型评价,我们只要对所选用的若干机构分别按上述步骤算出各机构的模糊决策集  $B_1^i, B_2^i, \dots, B_n^i$ , 然后综合评价它们的优劣,选择最佳机构。

对于由若干个机构组成的机械运动系统方案,亦可根据以上方法,先求出此机械运动系统方案中各机构的模糊决策集  $B_1, B_2, \dots, B_n$ ; 然后确定各机构的综合权重分配集  $A_{\sim}$ ; 最后计算此机械运动系统方案的模糊综合决策集  $B_{\sim}$ :

$$B_{\sim} = A_{\sim} \circ R_{\sim} = A_{\sim} \begin{bmatrix} B_{\sim 1} \\ B_{\sim 2} \\ \vdots \\ B_{\sim n} \end{bmatrix}$$

其中  $R_{\sim}$  可用各机构的模糊决策集叠加而成,其运算方法取  $M(\bullet, \oplus)$ 。

为了要对多个机械运动系统方案进行模糊综合评价,可将分别求出各方案的模糊综合决策集  $B_{\sim}^1, B_{\sim}^2, \dots, B_{\sim}^n$ 。根据模糊综合决策集的评价结果,在各方案中选择最佳方案。

例如,某种机械运动系统方案由3个执行机构组成,它有两套方案,已知:

$$A_{\sim}^1 = (0.4, 0.3, 0.3)$$

$$R_{\sim}^1 = \begin{bmatrix} 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.2 & 0.1 \end{bmatrix}$$

$$A_{\sim}^2 = (0.35, 0.35, 0.3)$$

$$R_{\sim}^2 = \begin{bmatrix} 0.35 & 0.3 & 0.2 & 0.15 & 0 \\ 0.4 & 0.4 & 0.1 & 0.1 & 0 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

由此可求出模糊综合决策集  $B_{\sim}^1, B_{\sim}^2$ :

$$\begin{aligned}
 B_{\sim}^1 &= (0.4, 0.3, 0.3) \begin{bmatrix} 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.2 & 0.1 \end{bmatrix} \\
 &= (0.385, 0.255, 0.16, 0.13, 0.10)
 \end{aligned}$$

归一化后得:

$$B_{\sim}^{1*} = (0.374, 0.248, 0.155, 0.126, 0.097)$$

$$\begin{aligned}
 B_{\sim}^2 &= (0.35, 0.35, 0.3) \begin{bmatrix} 0.35 & 0.3 & 0.2 & 0.15 & 0 \\ 0.4 & 0.4 & 0.1 & 0.1 & 0 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \end{bmatrix} \\
 &= (0.3525, 0.335, 0.165, 0.1175, 0.03)
 \end{aligned}$$

归一化后得:

$$B_{\sim}^{2*} = (0.3525, 0.335, 0.165, 0.1175, 0.03)$$

由  $B_{\sim}^{1*}, B_{\sim}^{2*}$  的评价结果来看,方案 I 的“很好”、“好”、“较好”占 77.7%, 方案 II 的“很好”、“好”、“较好”占 85.25%。因此,应选择方案 I。

如果机械运动系统方案由更多的执行机构所组成,提出的机械运动系统方案数更多,那么可以按上法求出  $B_{\sim}^{1*}, B_{\sim}^{2*}, \dots, B_{\sim}^{n*}$  后,最终选定某一方案。

## 6 实例分析

### 6.1 系统工程评价法评价机械运动方案

为了使提花织物纹板轧制系统实现自动化,纹版冲孔机的第一个功能是削纸,它是将放在纸库内的纹板(它是一块长400mm、宽68mm、厚0.8mm的纸板)推出,送至由一对滚轮组成的纹版步进机构。削纸机构的削纸速度要求均匀,每次削纸要可靠不能卡纸或削空。图47.7-8表示这种机构的结构示意图。同时,还要

求机构尽量简单,便于加工制造,便于设计。

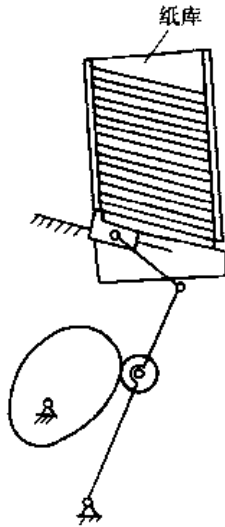


图 47.7-8 削纸机构

根据对削纸机构的要求,通过初步分析研究,可以采用以下 3 个方案:

- 1) 凸轮摇杆滑块机构(图 47.7-9);

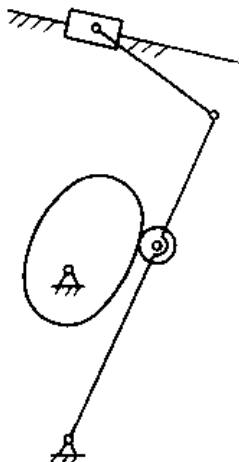


图 47.7-9 凸轮摇杆滑块机构

- 2) 牛头刨机构(图 47.7-10);

- 3) 斯蒂芬森机构(图 47.7-11)。

除了上述三种机构外,我们还可以通过创新和构思设计出其他型式的削纸机构。

根据削纸机构的工作特点、性能要求和应用场合等,采用表 47.7-2 所示的评价体系,它可以用图 47.7-12 来简单表示。

根据各评价指标相互关系,建立评价模型为:

$$H_A = \langle U_1(\cdot) \cdot U_2(\cdot) \cdot U_3(\cdot) \cdot U_4(\cdot) \cdot U_5 \rangle$$

式中  $U_1 = S_1 + S_2$ ;

$$U_2 = S_3 + S_4 + S_5 + S_6$$

$$U_3 = S_7 + S_8 + S_9 + S_{10}$$

$$U_4 = S_{11} + S_{12} + S_{13} + S_{14}$$

$$U_5 = S_{15} + S_{16} + S_{17}$$

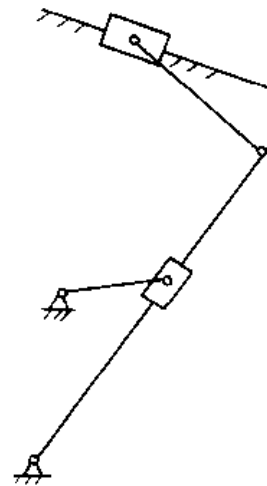


图 47.7-10 牛头刨机构

从上述表达式表示  $U_1, U_2, U_3, U_4, U_5$  各指标之间采用了乘法规则,而它们内部各子评价指标采用加法规则。

表 47.7-9 表示上述 3 个机构方案的评价指标体系、评价价值及计算结果。在表中所有指标值分为五个等级:“很好”、“好”、“较好”、“不太好”、“不好”,它们分别用 1, 0.75, 0.50, 0.25, 0 来表示。确定指标值最好征集有设计经验的设计人员意见,采取他们评定的指标值的平均值可以更趋合理,具体评估时不妨一试。

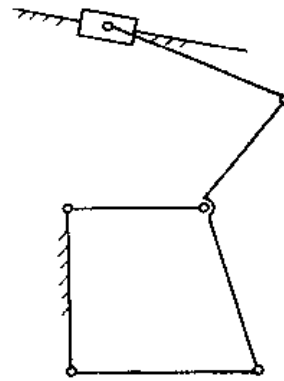


图 47.7-11 斯蒂芬森机构

根据表 47.7-9 表示的评价价值,用系统工程评价法可以算出各方案的  $H$  值,以  $H$  值的大小来排列 3 个机构方案的次序为:方案 I 最佳,方案 II 其次,方案 III 最差。在一般情况下宜选用方案 I。

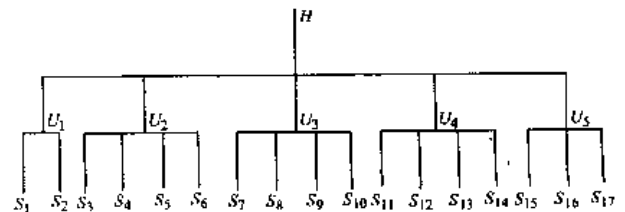


图 47.7-12 削纸机构评价体系

表 47.7-9 三种机构的评价体系、评价值和计算结果

评价指标		方案 I (凸轮摇杆滑块机构)	方案 II (牛头刨机构)	方案 III (斯蒂芬森机构)
U <sub>1</sub>	S <sub>1</sub>	1	0.75	0.75
	S <sub>2</sub>	0.75	0.75	0.75
U <sub>2</sub>	S <sub>3</sub>	0.75	0.75	0.75
	S <sub>4</sub>	0.75	0.75	0.75
	S <sub>5</sub>	0.75	0.75	0.75
	S <sub>6</sub>	0.75	0.75	0.75
U <sub>3</sub>	S <sub>7</sub>	1	0.50	0.50
	S <sub>8</sub>	0.50	0.75	0.75
	S <sub>9</sub>	0.50	0.75	0.75
	S <sub>10</sub>	1	1	1
U <sub>4</sub>	S <sub>11</sub>	0.50	0.75	0.50
	S <sub>12</sub>	0.50	0.75	0.75
	S <sub>13</sub>	1	0.75	0.75
	S <sub>14</sub>	0.75	0.75	0.75
U <sub>5</sub>	S <sub>15</sub>	0.75	0.50	0.50
	S <sub>16</sub>	0.75	0.75	0.75
	S <sub>17</sub>	0.75	0.75	0.50
方案的 H 值		89.32	81	78.875

6.2 模糊综合评价法评价机械运动系统方案

在冲压式蜂窝煤成型机运动系统方案设计过程中,可以看到它有 18 个方案可供选择。为了简化分析比较,我们将下列两个机械运动系统方案用模糊综合评价法来加以评估(表 47.7-10),如果有更多方案,亦可照此办理。

下面列出它的评价计算步骤:

表 47.7-10 蜂窝煤成型机评价的机械运动系统方案

蜂窝煤成型机的一大机构	机械运动方案 I	机械运动方案 II
冲头和脱模机构	对心曲柄滑块机构	六连杆冲压机构
扫屑刷机构	附加滑块摇杆机构	移动从动件固定凸轮机构
模筒转盘间歇运动机构	槽轮机构	凸轮式间歇运动机构

模糊决策集为

$$B_{-1}^i = A_{-1}^i \circ R_{-1}^i = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.5 & 0.2 & 0.2 & 0.1 & 0 \\ 0.5 & 0.2 & 0.1 & 0.2 & 0 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.2 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \end{bmatrix}$$

$$= (0.125 \oplus 0.1 \oplus 0.08 \oplus 0.08 \oplus 0.06, 0.05 \oplus 0.04 \oplus 0.04 \oplus 0.04 \oplus 0.045,$$

$$0.05 \oplus 0.02 \oplus 0.04 \oplus 0.04 \oplus 0.03, 0.025 \oplus 0.04 \oplus 0.02 \oplus 0.04 \oplus 0.015,$$

(1) 计算方案 I 中各机构的模糊决策集

1) 对心曲柄滑块机构

权数分配集  $A_{-1}^1 = (0.25, 0.2, 0.2, 0.2, 0.15)$

$$R_{-1}^1 = \begin{bmatrix} 0.5 & 0.2 & 0.2 & 0.1 & 0 \\ 0.5 & 0.2 & 0.1 & 0.2 & 0 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.2 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \end{bmatrix}$$

$$\begin{aligned} & 0 \oplus 0 \oplus 0.02 \oplus 0 \oplus 0) \\ & = (0.445, 0.215, 0.18, 0.14, 0.02) \end{aligned}$$

归一化后为

$$B_{-1}^{-1} = (0.445, 0.215, 0.18, 0.14, 0.02)$$

2) 附加滑块摇杆机构

$$\text{权数分配集 } A_{-2}^{-1} = (0.25, 0.2, 0.2, 0.2, 0.15)$$

$$\text{评价矩阵为 } R_{-2}^{-1} = \begin{bmatrix} 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0.2 & 0 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.5 & 0.3 & 0.1 & 0.1 & 0 \end{bmatrix}$$

模糊决策集为

$$\begin{aligned} B_{-2}^{-1} &= A_{-2}^{-1} \odot R_{-2}^{-1} = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0.2 & 0 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.5 & 0.3 & 0.1 & 0.1 & 0 \end{bmatrix} \\ &= (0.075 \oplus 0.06 \oplus 0.06 \oplus 0.08 \oplus 0.075, 0.075 \oplus 0.06 \oplus 0.06 \oplus 0.06 \oplus 0.045, \\ & \quad 0.05 \oplus 0.04 \oplus 0.04 \oplus 0.04 \oplus 0.015, 0.025 \oplus 0.04 \oplus 0.02 \oplus 0.02 \oplus 0.015, \\ & \quad 0.025 \oplus 0 \oplus 0.02 \oplus 0 \oplus 0) \\ &= (0.35, 0.3, 0.185, 0.12, 0.045) \end{aligned}$$

归一化后为

$$B_{-2}^{-1} = (0.35, 0.3, 0.185, 0.12, 0.045)$$

3) 槽轮机构

$$\text{权数分配集 } A_{-3}^{-1} = (0.25, 0.2, 0.2, 0.2, 0.15)$$

$$\text{评价矩阵为 } R_{-3}^{-1} = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0 \end{bmatrix}$$

模糊决策集为

$$\begin{aligned} B_{-3}^{-1} &= A_{-3}^{-1} \odot R_{-3}^{-1} = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0 \end{bmatrix} \\ &= (0.1 \oplus 0.08 \oplus 0.06 \oplus 0.08 \oplus 0.045, 0.05 \oplus 0.06 \oplus 0.04 \oplus 0.06 \oplus 0.045, \\ & \quad 0.05 \oplus 0.02 \oplus 0.04 \oplus 0.02 \oplus 0.045, 0.025 \oplus 0.02 \oplus 0.04 \oplus 0.02 \oplus 0.015, \\ & \quad 0.025 \oplus 0.02 \oplus 0.02 \oplus 0.02 \oplus 0) \\ &= (0.365, 0.255, 0.175, 0.12, 0.085) \end{aligned}$$

归一化后为

$$B_{-3}^{-1} = (0.365, 0.255, 0.175, 0.12, 0.085)$$

(2) 计算方案 I 中各机构的模糊决策集



## 1) 六连杆冲压机构

权数分配集  $A_{\sim 1}^1 = (0.25, 0.2, 0.2, 0.2, 0.15)$

$$\text{评价矩阵为 } R_{\sim 1}^1 = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.3 & 0 \end{bmatrix}$$

模糊决策集为

$$B_{\sim 1}^1 = A_{\sim 1}^1 \odot R_{\sim 1}^1 = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.3 & 0 \end{bmatrix}$$

$$= (0.1 \oplus 0.08 \oplus 0.08 \oplus 0.06 \oplus 0.045, 0.075 \oplus 0.04 \oplus 0.04 \oplus 0.04 \oplus 0.03, \\ 0.05 \oplus 0.04 \oplus 0.04 \oplus 0.04 \oplus 0.03, 0.025 \oplus 0.02 \oplus 0.02 \oplus 0.04 \oplus 0.045, \\ 0 \oplus 0.02 \oplus 0.02 \oplus 0.02 \oplus 0)$$

$$\therefore (0.365, 0.225, 0.2, 0.15, 0.06)$$

归一化后为

$$B_{\sim 1}^{*1} = (0.365, 0.225, 0.2, 0.15, 0.06)$$

## 2) 移动从动件固定凸轮机构

权数分配集  $A_{\sim 2}^1 = (0.25, 0.2, 0.2, 0.2, 0.15)$

$$\text{评价矩阵为 } R_{\sim 2}^1 = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.3 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

模糊决策集为

$$B_{\sim 2}^1 = A_{\sim 2}^1 \odot R_{\sim 2}^1 = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.3 & 0.1 & 0.1 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

$$= (0.1 \oplus 0.04 \oplus 0.04 \oplus 0.06 \oplus 0.06, 0.075 \oplus 0.04 \oplus 0.04 \oplus 0.04 \oplus 0.045, \\ 0.05 \oplus 0.06 \oplus 0.06 \oplus 0.06 \oplus 0.015, 0.025 \oplus 0.04 \oplus 0.04 \oplus 0.02 \oplus 0.015, \\ 0 \oplus 0.02 \oplus 0.02 \oplus 0.02 \oplus 0.015)$$

$$\therefore (0.3, 0.24, 0.245, 0.14, 0.075)$$

归一化后为

$$B_{\sim 2}^{*1} = (0.3, 0.24, 0.245, 0.14, 0.075)$$

## 3) 凸轮式间歇运动机构

权数分配集  $A_{\sim 3}^1 = (0.25, 0.2, 0.2, 0.2, 0.15)$

$$\text{评价矩阵为 } R_{\sim 3}^1 = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.3 & 0.2 & 0.1 & 0.2 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.2 & 0.1 \end{bmatrix}$$

模糊决策集为

$$\begin{aligned}
 B_{\sim 3}^I &= A_{\sim 2}^I \circ R_{\sim 2}^I = (0.25, 0.2, 0.2, 0.2, 0.15) \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.3 & 0.2 & 0.1 & 0.2 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.2 & 0.1 \end{bmatrix} \\
 &= (0.1 \oplus 0.06 \oplus 0.08 \oplus 0.06 \oplus 0.03, 0.05 \oplus 0.04 \oplus 0.06 \oplus 0.04 \oplus 0.045, \\
 &\quad 0.05 \oplus 0.04 \oplus 0.04 \oplus 0.02 \oplus 0.03, 0.025 \oplus 0.04 \oplus 0.02 \oplus 0.04 \oplus 0.03, \\
 &\quad 0.025 \oplus 0.02 \oplus 0.04 \oplus 0.015) \\
 &= (0.33, 0.235, 0.18, 0.155, 0.1)
 \end{aligned}$$

归一化后为

$$B_{\sim 3}^{*I} = (0.33, 0.235, 0.18, 0.155, 0.1)$$

(3) 两机械运动方案的模糊综合评价

1) 方案 I

方案 I 3 个执行机构的权数分配集取

$$A_{\sim 3}^I = (0.4, 0.25, 0.35)$$

方案 I 的综合评价矩阵, 由前可得

$$R_{\sim 3}^I = \begin{bmatrix} B_{\sim 1}^I \\ B_{\sim 2}^I \\ B_{\sim 3}^I \end{bmatrix} = \begin{bmatrix} 0.445 & 0.215 & 0.18 & 0.14 & 0.02 \\ 0.35 & 0.3 & 0.185 & 0.12 & 0.045 \\ 0.365 & 0.255 & 0.175 & 0.12 & 0.085 \end{bmatrix}$$

它的模糊综合决策集为

$$\begin{aligned}
 B_{\sim 3}^I &= A_{\sim 3}^I \circ R_{\sim 3}^I = (0.4, 0.25, 0.35) \begin{bmatrix} 0.445 & 0.215 & 0.18 & 0.14 & 0.02 \\ 0.35 & 0.3 & 0.185 & 0.12 & 0.045 \\ 0.365 & 0.255 & 0.175 & 0.12 & 0.085 \end{bmatrix} \\
 &= (0.178 \oplus 0.0875 \oplus 0.127753, 0.086 \oplus 0.075 \oplus 0.08925, \\
 &\quad 0.072 \oplus 0.04625 \oplus 0.06125, 0.056 \oplus 0.03 \oplus 0.042, \\
 &\quad 0.008 \oplus 0.01125 \oplus 0.02975) \\
 &= (0.3933, 0.2503, 0.1795, 0.1279, 0.049)
 \end{aligned}$$

归一化后为

$$B_{\sim 3}^{*I} = (0.3933, 0.2503, 0.1795, 0.1279, 0.049)$$

2) 方案 II

方案 II 3 个执行机构的权数分配集取

$$A_{\sim 3}^{II} = (0.4, 0.25, 0.35)$$

方案 II 的综合评价矩阵, 由前可得

$$R_{\sim 3}^{II} = \begin{bmatrix} B_{\sim 1}^{II} \\ B_{\sim 2}^{II} \\ B_{\sim 3}^{II} \end{bmatrix} = \begin{bmatrix} 0.365 & 0.225 & 0.2 & 0.15 & 0.06 \\ 0.3 & 0.24 & 0.245 & 0.14 & 0.075 \\ 0.33 & 0.235 & 0.18 & 0.155 & 0.1 \end{bmatrix}$$

它的模糊综合决策集为

$$B_{\sim 3}^{II} = A_{\sim 3}^{II} \circ R_{\sim 3}^{II} = (0.4, 0.25, 0.35) \begin{bmatrix} 0.365 & 0.225 & 0.2 & 0.15 & 0.06 \\ 0.3 & 0.24 & 0.245 & 0.14 & 0.075 \\ 0.33 & 0.235 & 0.18 & 0.155 & 0.1 \end{bmatrix}$$

$$\begin{aligned}
 &= (0.146 \oplus 0.075 \oplus 0.1155, 0.09 \oplus 0.06 \oplus 0.08225, \\
 &\quad 0.08 \oplus 0.06125 \oplus 0.063, 0.06 \oplus 0.035 \oplus 0.05425, \\
 &\quad 0.024 \oplus 0.01875 \oplus 0.035) \\
 &= (0.3365, 0.2323, 0.2042, 0.1493, 0.0777)
 \end{aligned}$$

归一化后为

$$\tilde{B}_{\text{总}}^{-1} = (0.3365, 0.2323, 0.2042, 0.1493, 0.0777)$$

#### (4) 机械运动系统方案的评估与选择

从上述计算所得  $\tilde{B}_{\text{总}}^{-1}$ 、 $\tilde{B}_{\text{总}}^{-1}$  来看, 方案 I 的“很好”、“好”、“较好”占 82.31%, 方案 II 的“很好”、“好”、“较好”占 77.30%。因此, 一般情况下应选择方案 I。

### 参 考 文 献

- 1 邹慧君编著. 机构系统设计. 上海: 上海科学技术出版社, 1996
- 2 邹慧君主编. 机械运动方案设计手册. 上海: 上海交通大学出版社, 1994
- 3 邹慧君主编. 机械原理课程设计手册. 北京: 高等教育出版社, 1998
- 4 邹慧君主编. 机械原理. 北京: 高等教育出版社, 1999
- 5 邹珊刚等编著. 系统科学. 上海: 上海人民出版社, 1987
- 6 江应洛主编. 系统工程理论方法与应用. 北京: 高等教育出版社, 1992
- 7 张延欣等编著. 系统工程学. 北京: 气象出版社, 1997
- 8 高富强等编著. 知识与经济信息化. 北京: 社会科学文献出版社, 1998
- 9 胡胜海主编. 机械系统设计. 哈尔滨: 哈尔滨工程大学出版社, 1997
- 10 苗东升著. 系统科学精要. 北京: 中国人民大学出版社, 1998
- 11 黄纯颖主编. 设计方法学. 北京: 机械工业出版社, 1992
- 12 廖林清等编著. 机械设计方法学. 重庆: 重庆大学出版社, 1996
- 13 威廉·卡尔文著. 大脑如何思维. 杨雄里等译. 上海: 上海科学技术出版社, 1996
- 14 赵惠田主编. 发明创造技法. 北京: 科学普及出版社, 1988
- 15 G. 帕尔, W. 拜茨著. 工程设计学. 北京: 机械工业出版社, 1992
- 16 黄靖远等主编. 机械设计学. 北京: 机械工业出版社, 1991
- 17 Pahl G., Beitz W. Engineering Design, London: The Design Council, 1984
- 18 Pahl G., Beitz W. 著. 工程设计学习与实践手册. 张直明等译. 北京: 机械工业出版社, 1992
- 19 M. J. French. Conceptual Design for Engineers (3<sup>rd</sup> edition) Springer-Verlag, 1999
- 20 H. S. Yan, Creative Design of Mechanical Devices, Springer Verlag, 1998
- 21 邹慧君等. 机电一体化系统概念设计的基本原理. 机械设计与研究, 1999(3)
- 22 邹慧君等. 机械产品概念设计及其方法综述. 机械设计与研究, 1998(2)
- 23 李学荣著. 新机器机构的创造发明—机构综合. 重庆: 重庆出版社, 1988
- 24 楼鸿棣, 邹慧君主编. 高等机械原理. 北京: 高等教育出版社, 1990
- 25 寺野寿朗著. 机械系统设计. 姜文炳译. 北京: 机械工业出版社, 1983
- 26 邹慧君等. 广义机构及其应用前景. 机械设计与研究. 2000年增刊
- 27 张启先, 张玉茹. 我国机械学研究的新进展与展望. 北京: 机械工程学报, 1996, 32(4): 1-4
- 28 Chiou S. J., Kota S.. Automated Conceptual Design of Mechanisms. Mechanism and Machine Design. 1999(34): 467-495
- 29 Qian L., Gero J. S. Function-behavior-structure Paths and their Role in Analogy-based Design. AIEDAM, 1996, 10(4): 289-312
- 30 邹慧君等. 机构学的研究现状、发展趋势和应用前景. 北京: 机械工程学报, 1999, 35(5): 1-4
- 31 Sharpe J. (Ed.), AI System Support for Conceptual Design. Springer, 1996
- 32 Sturges R. H., O'Shaughnessy K., Kilani M. I., Computational Model for Conceptual Design Based on Extended Function Logic, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 1996, 10(4): PP. 255-274
- 33 Qian L., Gero J. S. Function-behavior-structure Paths and their Role in Analogy-based Design. AIEDAM, 1996, 10(4): 289-312

- 34 Umeda Y., Takeda H., Tomiyama T., Yashikawa H., Function, Behavior, and Structure, In Application of Artificial Intelligence in Engineering, Gero J. S., Ed., Southhampton and Berlin: Computational Mechanics Publications and Springer-Verlag, 1990; pp. 177-193
- 35 朱崇贤. 工业设计系列讲座(七): 设计与市场. 机械设计与研究, 1994, (1): 44-47
- 36 荆冰彬等. 市场需求及其对产品的影响. 机械设计与研究, 1998, (1): 15-17
- 37 荆冰彬等. 基于市场分析的商品化产品设计目标决策研究. 中国机械工程, 1999, 10(1): 42-46
- 38 朱成钢. 市场学. 立信会计图书用品社, 1989
- 39 张春林等. 机械创新设计. 北京: 机械工业出版社, 1999.
- 40 黄纯颖主编. 机械创新设计. 北京: 高等教育出版社, 2000
- 41 邹慧君等编著. 机械系统概念设计. 北京: 机械工业出版社, 2003
- 42 邹慧君编著. 机械系统设计原理. 北京: 科学出版社, 2003

# 第 48 篇 反求设计与 快速成形制造技术

主 编 金 焯  
编写人 金 焯

## 概述

人类进入 20 世纪 90 年代以来,以计算机和信息处理技术为标志的科技飞速发展,对社会的生产和生活产生着深刻的影响,尤其对创造社会财富占 80% 的制造业面临着巨大的挑战,因为,人们消费的个性化和市场竞争的全球化,促使产品更新换代的周期愈来愈短,导致市场的快速多变,迫使企业必须寻找适合市场需求的新产品快速上市,占领市场,为此必须更新产品开发技术和创造新的产品设计、生产组合方式,淘汰大批量、单一产品的生产方式和落后的传统产品开发流程。为了适应制造领域的这种新形势,人们从生产的组织形式和高新技术的开发和应用都进行了广泛的研究,因而提出了诸如 CIMS、JIT、精良生产、并行工程、敏捷制造、虚拟企业等一系列新的现代生产组织和管理的哲理和思想,同时与此相应的一大批新的高新技术也不断涌现,其中快速反求设计和快速成形制造就是近十几年来在制造业中出现的具有里程碑的设计和

制造技术的变革。所谓反求设计,从广义的角度,它是 以物理对象(模型、实物、样件),软件(程序和技术文档),影像(图片、影视资料)为对象,研究对象的形态特征、工作原理、技术方案、功能、结构、材料、可靠性等等的一种技术,通常称反求工程技术。在机械工程领域,一般是指对物理对象的反求。快速原型制造是利用三维几何模型,通过分层和材料堆积直接制造物理原型的一种技术。这两种技术的有机结合为创新产品设计、提高新产品的的设计质量、新产品设计的一次成功率、响应快速多变的 市场提供了一种有效、快捷的工具。对一个企业的经营者来讲,最为重要的是要求有快速的新产品设计和开发能力去响应市场,用新产品占领市场,而快速反求设计和快速原型制造技术正是顺应了这种潮流的“宠儿”。用反求工程技术和快速成形制造应用于产品设计,已成为支持产品快速设计,创新设计和快速制造的重要支撑工具,这一技术在发达国家从 20 世纪 90 年代初开始应用于产品的设计和制造,而国内在近 4~5 年也开始逐步应用这一技术。

# 第 1 章 反求工程技术

## 1 反求工程的概念和流程

### 1.1 广义反求工程

反求工程(Reverse Engineering)这一术语起源于 20 世纪 60 年代,从通俗的意义上说,它是对现已存在对象的复制。现有对象可以是有形的物体,也可以是无形的软件、影像资料、工艺、材料成分和配方等等。反求是一个过程,通过这个过程获取对象的形态、工作原理、技术方案、功能、结构等等。对反求的目的本身,从法律的角度有很多问题有待研究,但这里我们仅从技术的角度来审视和研究。反求工程类似于反向推理,属于逆向思维体系。就机械工程领域的反求设计,它以设计方法学为指导,以现代设计理论、方法、技术为基础,运用各种专业人员的工程设计经验、知识和创新思维,对造型设计的油腻模型或已有的产品进行解剖、分析、重构和再创造,在工程设计领域,它具有独特的内涵,可以说它是对设计的设计。

反求工程技术是测量技术、数据处理技术、图形处理技术和加工技术相结合的一门综合性技术。随着计算机技术的飞速发展和上述单元技术的渐趋成熟,近年来在新产品设计开发中愈来愈多的被得到应用,成为产品快速开发的一条捷径。在产品开发过程中需要

以实物(样件)作为设计依据或参考模型或作为最终验证依据时尤其需要应用该项技术,所以在汽车、摩托车的外形覆盖件和内装饰件的设计,家电产品外形设计、艺术品复制中由于表面的形状复杂,很难用 CAD 的命令来构形时对反求工程技术的应用需求尤为迫切。

### 1.2 制造业反求工程的一般流程

反求工程的过程大致分为:首先由数据采集设备获取样件表面(有时需要内腔)数据,其次输入专门的数据处理软件或带有数据处理能力的三维 CAD 软件进行前处理,然后进行曲面和三维实体重构,在计算机上复现实物样件的几何形状,并在此基础上进行修改或重设计,最后对重构的对象进行实物制造。该方法主要用于对难以用现有 CAD 软件建模的曲面形状或未知设计方法的构件形状进行三维重构和再设计。其主要的流程如图 48-1-1。

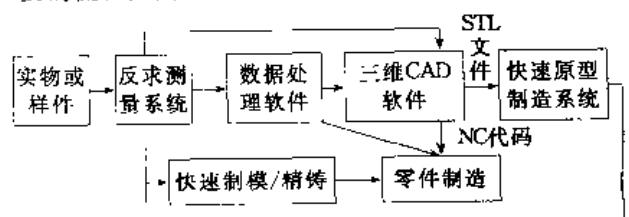


图 48.1-1 反求工程的一般流程

## 2 反求工程所涉及的主要技术

从数据采集到 CAD 模型的建立是反求工程最核心的部分,关键的技术都体现在其过程中,这个过程的

流程如图 48.1-2。

在这个流程中除根据不同样件对象和软硬件资源,规划合理的技术流程外,主要涵盖的技术有:

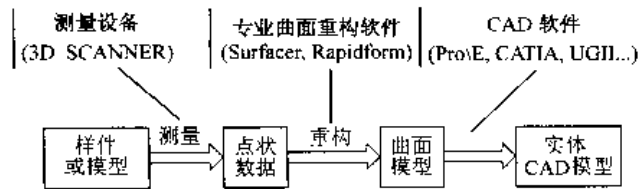


图 48.1-2 反求过程流程

### 2.1 数据采集技术

数据采集技术主要表现在数据测量设备的原理和方法上。目前可应用于自由曲面测量的方法和设备有许多种,这些测量方法和设备中,有的在自由曲面反求技术出现之前就已在工业界得到广泛应用,如 CMM、CT、MRI 等,有些是为了解决自由曲面测量而专门开发的方法和仪器,在现有自由曲面测量的测量仪器中,其原理和方法不同,因而各具优缺点,每种方法都有其侧重的工程应用领域和背景,这些测量方法的基本原理分别简要介绍如下。

#### (1) 坐标测量法(CMM)

在自由曲面的测量技术中,目前使用最为广泛和最为可靠的方法是坐标测量机(CMM)。这种测量方法,通常需要对被测对象进行分析,准备测量路径的规划,才能测得曲面的特征值。传统的测量机多采用触发式接触测量头,每一次获取自由曲面上一点的 X、Y、Z 坐标值。这种测量速度极慢,且难以测得较全面的曲面信息。20 世纪 90 年代初,英国 Renishaw 公司和意大利 DEA 公司等著名的坐标测量机生产厂,先后研制出了三维力一位移传感的扫描测量头,该测量头可以在工件上滑动测量,连续获取表面的坐标信息,其扫描速度最高可达 8m/min,数字化速度最高可达 500 点/s,精度也可达 30 $\mu$ m,由于这种扫描测头价格极为昂贵,目前还没有在坐标测量机上广泛采用,这种扫描测量头的固有缺陷在于它的扫描数字化速度受到机械限制,而且对不可触及的表面(如软表面、精密的光滑表面等)是无法测量的。

#### (2) 三角法

1) 单光点 单光点三角法是光学三维面形测量中最为常见的一种方法,它的基本原理是光束以一定角度投射在被测物体上,在另一方向上用 CCD 摄像机接受散射光的图像,利用三角几何关系来计算被测目标的深度信息。这种单光点的三角法已被广泛研究。由于这种方法一次只能获得一点的信息,使用中一般采

用整个测头的机械运动来获取整个物体上的三维坐标信息。这种方法类似于接触测量中的扫描测量头,它的结构简单,体积小,易于实现较高的测量精度,但在测量速度上仍显不足,目前这种测头一般是安装在 CMM 上,多用于小件物体或电路板检测等方面,它的测量分辨率最高可达 0.005mm。

2) 单光条 为了提高测量的速度,线光源投射的方法也已被广泛地采用,这种线光条是使用结构光测量中最简单的一种,也称为光刀法,早在 20 世纪 70 年代中期就提出了使用线光条提取三维信息。20 世纪 80 年代初,用激光做光源,并采用摄像机摄取结构光条的三维测量已被采用。进入 20 世纪 80 年代末 90 年代初,一种基于线结构光的三维型面测量方法已日趋成熟。出现了一些实用的产品。如英国 3D SCANAER 公司的 REPLECA 系列反求工程系统就是采用线结构光的方法获取物体表面的三维形貌信息,它的测量精度可达 50 $\mu$ m。法国的 KREON INDUSTRIE 公司 LASER SENSOR KL 系测头与软件组成的三维激光扫描测量系统,测量精度可达到 10~30 $\mu$ m。这种线结构光的测量方法具有单光点三角测量法的所有优点,同时又可实现高密度、高精度的表面面型测量,是目前较为实用可靠的一种三维非接触测量的方法。因此它也是目前自由曲面反求工程技术中三维面形信息快速获取的主要方法,这种线光条三维曲面测量的研究主要体现在精度和速度的提高方法上,主要集中在光条的产生与提取以及系统的物、像坐标的标定等方面。

3) 多光条 与单光条的结构光相类似,多光条的结构光测量法也是为了提高测量的速度,多光条的结构光测量法是一种全场测量的方法,它一次可以得到整个被测物体上一个区域的三维面形信息,目前多用光栅达到多光条测量,当光栅投影到物体表面时,周期性光栅的相位就受到物体表面高度轮廓的调制。形成变形光栅,变形光栅虽然是 2D 图像,但携带有 3D 信息。这种 3D 信息的提取方法也就引出了多种解调的手段。最简单的是基于三角几何关系的直接计算法,此

外,还有莫尔拓扑法,傅里叶变换轮廓法,相位步进解调法等等,以及上面这些方法中几种方法的综合。最近又有人提出了采用计算机产生模拟接收光栅的解调方法,多光条的结构光测量法具有比单光点扫描法或线结构光方法更高的测量速度,但其结构比前两者复杂得多,而且三维信息的提取方法上也要复杂一些,这种方法多用于对小型物体的一次测量,它的测量精度最高达0.01mm。

### (3) 图像法

这种方法是基于图像的三维信息获取,它是利用图像大小、明暗、纹理、光流等信息或利用立体视觉法求出三维信息。

1) 立体视觉法 这种方法类似于人眼看物体的方法,这种方法中用单摄像机的一般是用于对平面零件的检测,绝大多数是采用两个摄像机摄取两幅图像,从而通过两幅图像找出对应点,通过这些特征点匹配,得到物体表面某些位置上的点与相机光学中心之间的距离。为了得到比较稠密的表面点的距离数据从而得到距离图像,需要用内插方法,为了弥补两摄像机的不足,有些系统采用了三个相机、甚至四个相机的测量系统,这种立体视觉的方法一般情况下精度不高,它的分辨率一般是在毫米数量级,其最高也只有0.1mm,但该方法具有快速的信息获取能力,可实现动态测量,因而,这种方法一般用于机器人视觉和物体特征识别方面,目前,这种方法一般不用于反求工程中的测量。

2) 聚焦法 聚焦法是通过像距的调节,使聚焦正确亦即使物体轮廓的图像十分锋利,由此可获得物距,这就是从聚焦到距离估计的原理。所谓正确的聚焦是指由聚焦的锋利性来度量的,使物体轮廓图像达到锋利的过程,从聚焦得到的距离估计不需预先的图像分割,这是一个很大的优点,它的局限性是对灰度平坦的区域测距精度不高,另外聚焦法也和单点的三角法相同是单光点的测量,因而也需通过坐标测量机或数控铣床作为坐标移动来扫描测量工作,这种聚焦法一般可以达到较高的精度,分辨率一般可达1~10 $\mu$ m,但测量速度较慢,这种方法多用于测量精度高但对测量点密度要求不大的场合。

### (4) 断层测量法

近年来医用和工业界对诊断和测量的需求,发展起来多种断层扫描方法,这些方法同样可以用于自由曲面反求的测量中,这些方法主要有用超声波的数学方法,采用CT(Computed Tomography)或MRI(Magnetic Resonance Imaging)的断层扫描测量法以及CGI(Capture Geometrical Inside)的照相测量法,这些方法的一个共性就是可以得到复杂物体的内外表面信息,不受物体形状影响。

1) CT测量法 计算机断层扫描(CT)技术最具代表的是基于X射线的CT扫描机,它是以测量物体对X射线的衰减系数为基础,用数学方法经过电子计算机处理而重建断层图象。这种方法最早是用于医学领域,目前已开始用于工业领域,特别是对中空物体的无损三维测量。这种方法是目前最先进的非接触式检测方法,它可以对具有复杂内部结构的物体进行测量,但它存在空间分辨率低,获取数据需较长的积分时间,重建图像计算量大,造价高的问题。

2) MRI测量法 核磁共振(MRI)技术的理论基础是核物理学的磁共振理论,其基本原理是用磁场标定人体某层面的空间位置,然后用射频脉冲序列照射,当被激光的核在驰豫过程中自动恢复到静态场的平衡状态时,把吸收的能量发射出来,然后利用线圈来检测这种信号,由于这种技术具有深入物质内部,不破坏样品的优点。但这种方法的不足之处是其造价极为昂贵,而且对于金属物体不适用,这种方法目前主要还只是用于人体医学的三维测量。

3) CGI测量法 CGI是一种直接从样件到模型的方法,它是20世纪90年代初美国研制的一种测量方法,可对任意复杂样件的内外表面进行一次测量,它的具体做法是将工件内外进行实体填充后,用高速铣削的方法逐层铣削,然后再采用视觉照相的方法逐层测得每个截面的轮廓形状信息,该方法的最大优点是对任意复杂形状的样件进行快速反求测量,其不足之处是对样件进行破坏性测量,同时整个系统比较庞大复杂。

## 2.2 数据处理技术

上述各种不同的数据获取方法所获取的数据形式最终不外乎两种,即三维坐标形式和点阵图像形式,通过对这两类数据的处理,提取重构曲面必要的信息,这就是数据处理技术。

测量设备以及零件的表面特征等因素都可能使测量时产生一些噪点,去除噪点通常是反求工程中必不可少的过程。利用激光扫描仪测量得到的点通常数据量很大,通常称为“点云”数据,对这类数据,为了提高反求的效率,在精度允许的情况下,可以对点云进行删减。但另一方面,测量得到的数据也可能不完整,需要通过已测部分的数据来重构整个表面,这样就要涉及到表面的延伸、修补等技术。在某些情况下,要求建立整个物体的几何模型,但如果只采用一次测量或扫描只能得到物体一部分的信息,这时就需要移动物体或测量设备,并进行多次测量,这样就得到多片点云,将这多片点云拼合起来,就可以得到整个物体的表面信息。对于激光扫描仪测量获得的数据,一般噪点比较



少,主要涉及到的预处理工作包括多片点云拼合、重采样、点云法矢调整、三角剖分和点云集的划分。

(1) 多片点云数据的拼合

多片点云数据的拼合主要采用两种方式:辅助球方式和自动拼合方式。辅助球方式是在需要测量的物体上粘结若干个标准的球,需要拼合的两片点云中必须保证有三个球同时被测量到,这样就可以利用这三个球来对两片点云进行拼合。该方法虽然拼合过程简单,但是前期准备工作(如粘结剂预热等)比较费时,测量的效率会大大降低。因此采用自动拼合的方法实现多片点云的拼合是发展的方向。

多片点云拼合中的一个主要的问题是拼合精度问题,多片点云拼合的实质是寻找将一片点云的信息与另一片点云的信息相关联的拼合矩阵(旋转或平移),使物体的所有信息可以拼合到一起。转换矩阵的求解一般利用迭代的方法(Iterative Closet Point,简称ICP)得到,首先选取一个初始拼合矩阵,建立拼合误差目标函数,然后利用迭代方法求解,得到最优的拼合矩阵,将两片点云拼合到一起。目前的多片点云自动拼合方法存在一个问题,即拼合精度受初始点的影响较大。如果初始点选择不当,一方面会影响拼合的速度,另一方面有可能陷入局部最优,达不到给定的精度。

这里介绍一种基于弱特征的点云数据拼合技术。所谓弱特征就是点云的特征边界点的集合。这些特征点集可能是直线、二次曲线或自由曲线,但我们在拼合时不需要知道线的特征,只需要将它们看作点集就可以了,因此我们称之为弱特征。首先,介绍在点云中弱特征的提取方法,然后给出基于弱特征的点云数据拼合技术。

对于单个视图,可利用基于图像法提取点云的弱特征。在二维图像的边缘检测算法中,通常采用边缘检测算子与原图像卷积来得到一幅边缘幅度图像,再利用这幅边缘幅度图像提取二维图像的边缘。边缘检测算子是用来检查二维图像上每个象素的领域并对灰度变化率进行量化,通常也包括方向的确定,大多数算子采用方向导数掩模,例如 Roberts 算子、Sobel 算子、Prewitt 算子等。这里采用 Prewitt 方向掩模来求解两个方向导数,然后利用方向导数求解法矢和离散曲面上各点的曲率值,利用上述曲率值就可以用来提取离散点云曲面的边界点集。

对于采用激光扫描仪测量得到的点云数据,可以看成是一幅深度图像或一个离散二次曲面,因此,可表示为:

$$X(u,v) = [u, v, f(u,v)]^T \quad (48.1-1)$$

式中  $X(u,v)$ ——二次曲面, $u,v$ 为二次曲面的两个方向参数;

$f(u,v)$ ——深度值。

则两个方向导数分别为:

$$X_u = [1 \ 0 \ f_u]^T, \quad X_v = [0 \ 1 \ f_v]^T \quad (48.1-2)$$

单位法矢为

$$N = \frac{X_u \times X_v}{|X_u \times X_v|} = \frac{[-f_u \ -f_v \ 1]^T}{\sqrt{1+f_u^2+f_v^2}} \quad (48.1-3)$$

方向导数  $f_u, f_v$  可以分别用 Prewitt 算子的两个主向掩模得到。

沿  $u, v$  两个方向的掩模算子分别为  $D_u, D_v$ :

1	1	1
0	0	0
-1	-1	-1

1	0	-1
1	0	-1
1	0	-1

则:

$$\begin{aligned} f_u &= D_u * f(u,v) \\ f_v &= D_v * f(u,v) \end{aligned} \quad (48.1-4)$$

式中  $*$ ——二维卷积。

这样,利用式(1-4),就可以求出曲面上任意一点的法矢。然后,进一步求解曲率,曲率可以看作是法矢沿给定方向的变化率。对于一个离散曲面,假设其中某一点  $p$  的法矢为  $N$ ,指定一方向向量  $(Q-P)$  为从点  $p$  指向它的领域上的一点  $q$ ,则在点  $p$  处沿该方向的曲率为:

$$k(p,q) = \frac{|N_q - N_p|}{|Q - P|} s(p,q) \quad (48.1-5)$$

式中的  $s(p,q)$  表示曲率的符号,具体为:

$$s(p,q) = \begin{cases} 1 & (N_p \times N_q) \cdot (N_p \times (Q - P)) > 0 \\ -1 & \text{其他} \end{cases} \quad (48.1-6)$$

利用式(1-5),可以估计出离散曲面任一点的 4 个基本曲率表示,假设点  $p$  的  $3 \times 3$  领域  $\Omega(p)$ ,各曲率表示如下:

① 最小曲率:  $k_{\min}(p) = k(p, q_1) = \min_{q \in \Omega(p)} k(p, q)$ ,

最小曲率方向:  $D_1 = Q_1 - P$ ;

② 最大曲率:  $k_{\max}(p) = k(p, q_2) = \max_{q \in \Omega(p)} k(p, q)$ ,

最大曲率方向:  $D_2 = Q_2 - P$ ;

③ 平均曲率:  $k_{\text{mean}}(p) = (k_{\min}(p) + k_{\max}(p))/2$

④ 高斯曲率:  $k_{\text{gauss}}(p) = k_{\min}(p) \times k_{\max}(p)$

利用上述曲率值就可以用来提取离散点云曲面的边界点集。

离散曲面的边界点可以定位为曲率极值点或零交叉点,在提取边界点时,可以首先设定一个阈值  $\gamma$ ,以曲面的平均曲率为对象,如果曲面的平均曲率大于阈值,则作为候选边界点。这样就可以得到一系列候选边

界点,实际边界点就在这些候选边界点内产生。

对这些候选边界点进行操作,实际的边界点为曲率极值点,即只要某一点的 2 个主曲率中任何一个沿对应的主方向上为极值,那么就可以将该点定义为边界点。如,采用近似计算的方法,选取一候选边界点  $p$ ,首先沿最小曲率方向  $D_1$ ,找到左右 2 个邻接点  $q_{1l}, q_{1r}$ ,如果  $k_{\min}(p) \leq k(q_{1l})$  并且  $k_{\min}(p) \leq k(q_{1r})$  ( $k(q_{1l})$ 、

$k(q_{1r})$  分别为点  $q_{1l}, q_{1r}$  沿最小曲率方向  $D_1$  的曲率值,那么该点就是实际的边界点;否则,再沿最大曲率方向  $D_2$ ,找到该点的左右两个邻接点  $q_{2l}, q_{2r}$ ,如果  $k_{\max}(p) \geq k(q_{2l})$  并且  $k_{\max}(p) \geq k(q_{2r})$ ,那么点  $p$  也可作为实际的边界点。对所有候选边界点进行上述操作,就可以得到所需要的全部边界点。具体流程如图 48.1-3 所示。

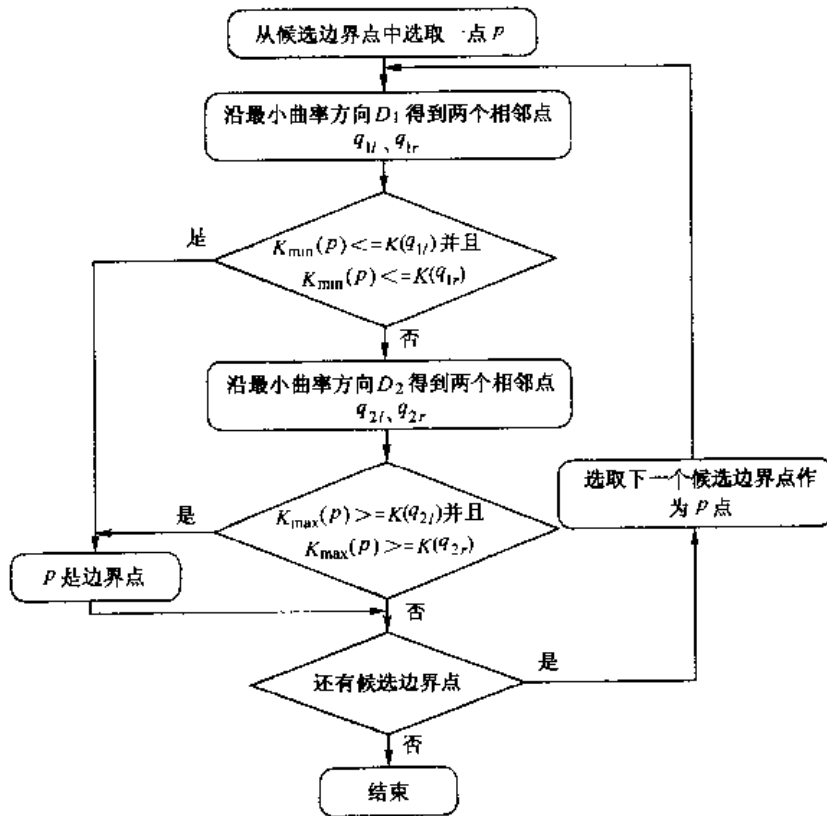


图 48.1-3 边界点自动提取流程

通过上述方法可以得到点云数据的边界点集,根据点集的连通情况将它们分类,相互连通的点集成为一个弱特征。在进行实际拼合时,只需要在拼合点云中选取三对弱特征就可以实现拼合。弱特征的拼合过程如下:

假定有两片待拼合的点云  $A\{p, i=1, 2, \dots, m\}$ ,  $B\{p, j=1, 2, \dots, n\}$ ,其中点云 A 固定不动,通过移动点云 B 来和点云 A 拼合到一起。在点云 A 中存在三个弱特征  $Fa1, Fa2, Fa3$ ,在点云 B 中存在对应的三个弱特征  $Fb1, Fb2, Fb3$ ,对应弱特征对内部的数据点的个数可以不相同。利用人机交互的方式,在这三对弱特征中分别选取对应的三个点作为初始值。

在  $Fa1, Fa2, Fa3$  中分别选取三点  $P_{a1}(x_{a1}, y_{a1}, z_{a1}), P_{a2}(x_{a2}, y_{a2}, z_{a2}), P_{a3}(x_{a3}, y_{a3}, z_{a3})$ ,在  $Fb1, Fb2, Fb3$  中选取同样位置的三点  $P_{b1}(x_{b1}, y_{b1}, z_{b1}), P_{b2}(x_{b2}, y_{b2}, z_{b2}), P_{b3}(x_{b3}, y_{b3}, z_{b3})$ ,利用这三点求解弱特征拼合

矩阵:

① 首先将点  $P_{b1}$  与点  $P_{a1}$  重合,其转换矩阵为  $T_1$ :

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ (x_{a1} - x_{b1}) & (y_{a1} - y_{b1}) & (z_{a1} - z_{b1}) & 1 \end{bmatrix}$$

② 然后将直线  $\overline{P_{a1}P_{b2}}$  与直线  $\overline{P_{a1}P_{a2}}$  重合,其变换矩阵为  $R_1$ ;设两直线之间的夹角为  $\theta_1$ ,过点  $P_{a1}$  且与两直线相垂直的直线为  $L_1$ ,利用绕任意轴旋转的变换公式就可以得到旋转变换矩阵  $R_1$ 。

③ 将  $\overline{P_{b1}P_{b3}}$  变换到与直线  $\overline{P_{a1}P_{a3}}$  重合,其变换矩阵为  $R_2$ 。

④ 得到点云拼合的初始变换矩阵  $M_0 = T_1 \cdot R_1 \cdot R_2$ 。

计算每一对弱特征的距离  $d$ ,  $d$  为一个弱特征中

的每一个点与对应弱特征中最近点之间的距离的平方和。

$$d = \sum_{i=1}^m (P_{a_i} - P_{b_k})^2$$

将三对弱特征的距离求和,建立目标函数:

$$f = \min \left( \sum_{i=1}^3 d_i \right)$$

利用迭代方法,可以得到弱特征的拼合矩阵。由于弱特征内部的点的个数较整个点云要小的多,因此拼合速度较快,而且精度也比仅用三个点进行拼合要高。然后将弱特征的拼合矩阵作为整个点云拼合的初始矩阵,这样不仅可以为点云拼合提供合适的初始条件,避免陷入局部最优,而且可以大大提高拼合速度。如果拼合点云内部的弱特征不足三对,也可以选取重复的弱特征,例如,如果只有一对弱特征,那么可以在这一对弱特征中选取不共线的三对数据点,然后进行特征拼合。

接下来利用 ICP 方法进行多片点云拼合,为了提高点云拼合的速度,需要先分别建立点云数据之间的拓扑邻域关系,可利用包围盒的方法实现。

将初始拼合后的点云 A 和点云 B 看成是一个统一的点云,记为点云  $C = A \cup B = \{c_k, k=1, 2, \dots, m+n\}$ 。先计算整个数据的包围盒大小,搜索全部点云数据,得到三个方向的最小、最大坐标值。 $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}, z_{\min}, z_{\max}\}$ ,然后根据要求将包围盒划分为若干个子立方体。设子立方体的边长为  $a$ ,则包围盒划分后三个方向的子立方体个数分别为:

$$m = \text{int}((x_{\max} - x_{\min})/a) + 1;$$

$$n = \text{int}((y_{\max} - y_{\min})/a) + 1;$$

$$r = \text{int}((z_{\max} - z_{\min})/a) + 1。$$

其中,  $\text{int}()$  为取整函数。然后根据每一个数据点的  $(x, y, z)$  坐标值将该数据点的序号追加到相应的立方体中去。不同的立方体根据其在三个方向的序号来识别。接下来利用初始拼合结果寻找两片点云之间的重叠点云区域。

① 从点云 A 中任取一点  $p_1(x_1, y_1, z_1)$ , 找到该点所在的立方体删格,在立方体内寻找是否有云 B 中的点,如果没有,则从点云 A 中选取新点,重复上述过程。

② 如果有,那么在该立方体中找到点云 B 中离  $p_1$  距离最近的一点  $p_1'(x_1', y_1', z_1')$ , 设定一个阈值  $v$ , 如果两点之间的距离大于  $v$ , 那么跳过点  $p_1$ , 重新执行 ①, 因为两点距离太远, 可以将该点排除在重叠点云区域之外。

③ 点  $p_1$  与从点  $p_1'$  构成一对拼合数据点。在点云 B 中的数据点  $p_1'$  的邻域选取与点  $p_1'$  不共线的两

点,与  $p_1'$  构成一个三角形  $\Delta_1$ , 计算点云 A 中的点  $p_1$  与三角形  $\Delta_1$  的距离  $d_1$ 。

设两片预拼合点云的第  $i$  对数据点之间的上述距离为  $d_i$ , 则:

$$d_i = A_i - (R \cdot B_i + T) \quad (48.1-7)$$

式中  $R$ ——旋转矩阵;

$T$ ——平移矩阵。

④ 迭代进行上述过程,得到两片待拼合点云的重叠点云区域与拼合误差,设点云 A 的重叠区域子点云  $A_s$ 、点云 B 的重叠区域子点云为  $B_s$ , 两片点云的拼合误差为  $e$ 。

$$e = \frac{1}{n} \sum_{i=1}^n w_i d_i^2 \quad (48.1-8)$$

式中  $n$ ——重叠区域拼合点云的对数;

$w_i$ ——第  $i$  个数据点相对应的权值。

利用迭代方法,计算拼合变换矩阵  $R$  和  $T$ , 使式(1-8)的值最小。迭代时,平移矩阵可以用如下的方法直接计算:

① 计算点云  $A_s, B_s$  的几何重心  $A_{sc}, B_{sc}$ :

$$A_{sc} = \frac{1}{n} \sum_{i=1}^n A_{si}, B_{sc} = \frac{1}{n} \sum_{i=1}^n B_{si}$$

② 计算平移矩阵  $T: T = A_{sc} - B_{sc}$ 。

利用迭代优化方法,可以计算出旋转矩阵,最终得到拼合后的两片点云。图 48.1-4 表示了一个汽车车头油泥模型的点云数据多视图拼合过程。图 48.1-4a, b 为两片需要拼合的点云,图 48.1-4c, d 为从两片点云中提取的弱特征点集,图 48.1-4e 为拼合后的点云,图 48.1-4f 为利用传统的 ICP 方法拼合的结果,可以看出在车轮部分点云拼合不理想。

(2) 散乱点云重新采样

由于各点云之间具有重叠的部分,因此拼合后的点云包含一些冗余的数据,需要将它们去除。利用重新采样的方法可以得到分布较为均匀的点云数据,这样的点云也比较适合下一步的三角剖分。

点云数据的删减方法有很多种,常用的有均匀采样法,弦长偏差法,空间采样法等。均匀采样法根据用户定义的采样率来进行采样,该方法要求点云数据采样前已经排好序。弦长偏差法根据给定的精度要求对点云数据进行采样,采样时需要给定两个参数:最大距离偏差值和两点之间的最大距离值。最大距离偏差确保欲删去的数据点与余下的数据点之间的偏差在给定范围内,以防止删去一些包含有用的几何信息的点,两点之间的最大距离值确保欲删去的数据点与余下的数据点之间的距离值在给定的范围内,以防止点云数据内部特别是几何形态变化不大的地方出现大的空隙。该方法使余下的点云较好地保留了原始点云的几何信

息。空间采样法利用给定的三维空间邻域大小采样点云数据,当某一点与点云数据之间的距离大于给定值

时,保留该点,否则删去该点。上述方法采样结果的好坏取决于点云数据的序列。

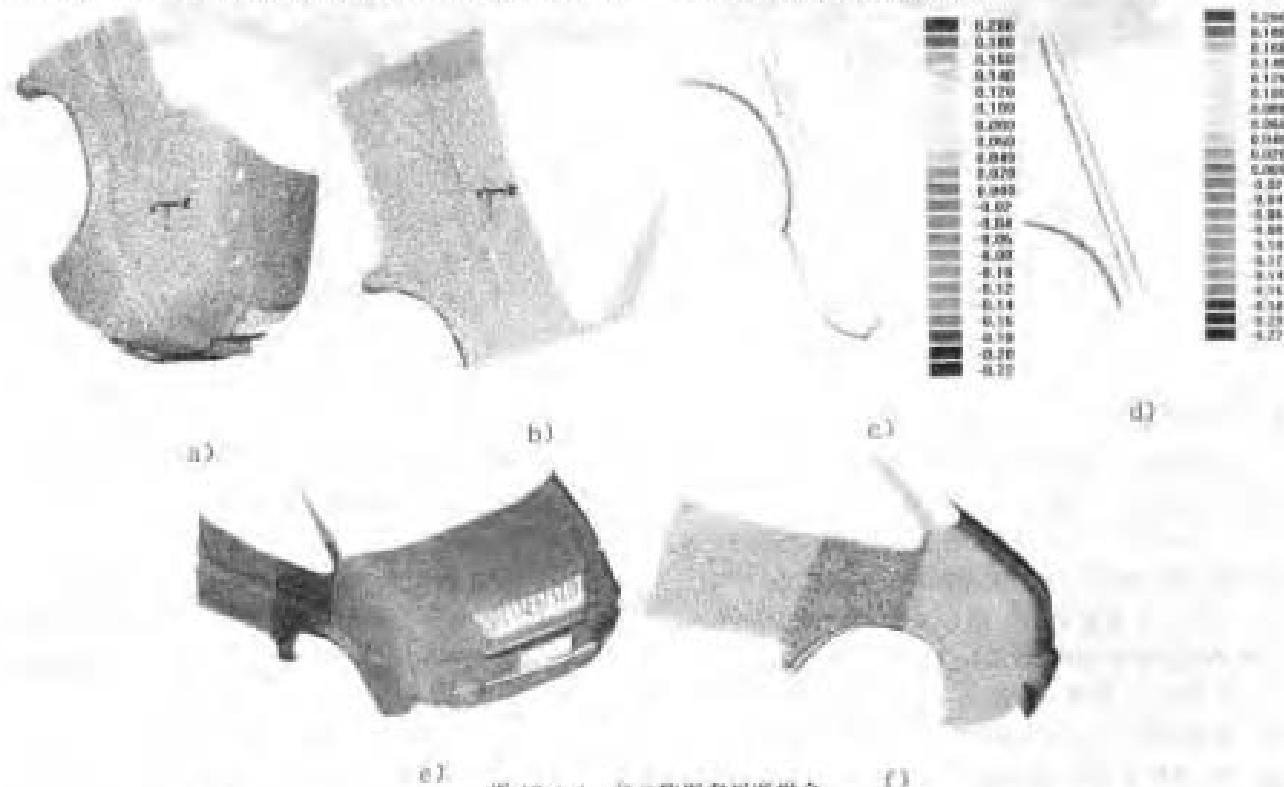


图 48-1-4 点云数据多视图融合

a)点云视图1 b)点云视图2 c)网格特征点集1 d)网格特征点集2 e)融合结果 f)ICP方法融合结果

以下介绍几种点云数据删减方法。

利用 3D 网格对点云数据进行删减的方法,先对点云进行局部三角剖分并计算法矢,生成初始网格,然后采用 8 叉树数据结构对网格细分,在每一个子网格中提取一个数据点,最终得到删减后的点云。该方法采样得到的点云数据可以较好地保留原始几何信息,且数据点分布较均匀,但过程较复杂,尤其是点云数据很密时,算法的效率大大下降,另外该方法对噪声比较敏感。

采用三维空间滤波器方法对点云进行删减:先计算全部点云数据的包围盒,然后对每一个数据点局部插值二次曲面,接着计算曲面方程的最大二阶偏导数,用来确定三角面片的最大边长和点云—三角面片之间的误差 $\epsilon$ ,根据精度要求计算子盒尺寸,最后利用子盒来删减点云。利用该方法采样后得到的数据点分布均匀,但容易丢失特征等几何信息。

采用改进的三维空间采样法实现对点云数据的采样,该方法过程简单,而且可以较好地保留原始点云的几何信息,具体方法是:

① 首先定义采样间距的大小,即每隔多少空间距离采样一个点。

② 将点云数据进行空间划分,形成若干个包含

点云的子立方体,子立方体的边长为采样间距。

③ 在每一个子立方体内,计算该子立方体内所有数据点的平均法矢,从该子立方体内选取法矢值与该平均法矢之间的夹角最小的点作为采样点,删去其余的点。因为该点能较好地反映该区域的几何形态。

图 48-1-5 表示的点云数据的重采样过程。图 48-1-5a 为原始点云,包含 86363 个数据点。图 48-1-5b 为点云数据法矢,图 48-1-5c 为重采样后得到的点云,采样间距为 3mm,采样后的数据点个数为 10910,图 48-1-5d 为另一重采样后得到的点云,采样间距为 4mm,采样后的数据点个数为 4803。

### (3) 点云法矢调整

在进行三角剖分之前,要求估计出点云中各点处的法矢,这样可以改善三角网格曲面的精度。具体过程是首先建立点云数据的  $k$  邻域,然后利用  $k$  邻域估计点云数据的法矢,最后进行三角剖分,建立三角网格曲面。

首先确定数据点的  $k$  邻域。先选定点云数据空间的包围盒,然后将包围盒进一步分割成许多小的立方体,各立方体首尾相接,因此每个立方体顶点就与 8 个立方体相邻,设有一数据点  $x_i$ ,根据其坐标值计算出离该点最近的栅格顶点位置,然后在与该顶点相邻的 8 个立方体中寻找距离点  $x_i$  最近的  $k$  个数据点,这  $k$  个

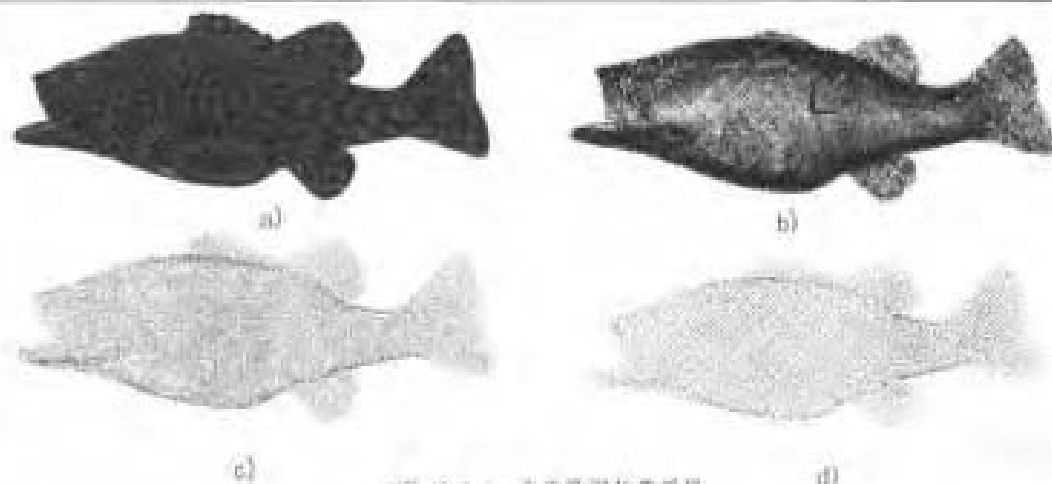


图 48-1-3 点云数据的重采样

数据点就构成了点  $x_i$  的  $k$  邻域。

若以每一个数据点为顶点,该点与相邻的  $k$  个邻接点分别连一条边,形成一个无向图,称为 Riemannian 图。Riemannian 图表示了点云数据的  $k$  邻域情况。通常图的存储采用邻接矩阵和邻接表的形式。当  $k$  远远小于数据点的个数  $n$  时(一般  $k$  取 20~30),通常采用邻接表的形式,但链表运算速度较矩阵慢,结合二者的优点,采用一种新的存储形式:双邻接矩阵形式。即利用两个  $n \times k$  矩阵  $M_1, M_2$  来存储 Riemannian 图。 $M_1$  为序号矩阵, $M_1$  的每一行用来存储与  $i$  号顶点相邻的  $k$  个邻接点的序号,并按与  $i$  号顶点的距离由小到大顺序排列。例如  $M_1(2,3)=10$ ,表示在 2 号点的  $k$  个邻接点中,第 3 个点是 10 号点,10 号点与 2 号点之间有一条边。 $M_2$  为权值矩阵, $M_2$  与  $M_1$  的各元素一一对应,用来存储与  $M_1$  相对应的边上的权值。权值可以是距离,也可以是其他值。例如,以距离作为权值,若

$$\begin{bmatrix} \sum_{i=1}^k (x_{i1} - O_x)^2 & \sum_{i=1}^k (x_{i1} - O_x)(y_{i1} - O_y) & \sum_{i=1}^k (x_{i1} - O_x)(z_{i1} - O_z) \\ \sum_{i=1}^k (y_{i1} - O_y)(x_{i1} - O_x) & \sum_{i=1}^k (y_{i1} - O_y)^2 & \sum_{i=1}^k (y_{i1} - O_y)(z_{i1} - O_z) \\ \sum_{i=1}^k (z_{i1} - O_z)(x_{i1} - O_x) & \sum_{i=1}^k (z_{i1} - O_z)(y_{i1} - O_y) & \sum_{i=1}^k (z_{i1} - O_z)^2 \end{bmatrix}$$

式中  $x_{i1}, y_{i1}, z_{i1}$  ——  $P_{i1}$  的三个坐标值;

$O_x, O_y, O_z$  —— 重心  $O$  的三个坐标值。

该矩阵为实对称矩阵,利用 Jacobian 方法就可以求解其特征值。设矩阵的最小特征值分别为  $\lambda_{min}$ ,对应的特征向量为  $V_{min}$ ,则最小二乘拟合平面的法矢为:

$$\{N\} = \{V_{min}\}$$

即最小二乘拟合平面的法矢可以看作是协方差矩阵的最小特征值对应的特征向量。

通常采用前面的方法得到的点云数据法矢的方向是不协调的,即有些数据点处的法矢指向曲面内侧,有些数据点处的法矢指向曲面外侧(图 48-1-6b)。因此,需要对上述法矢方向进行调整,使所有法矢都指向曲

面的同一侧。

法矢调整算法通常从点云中某一点开始,一般选取  $z$  坐标值最大的点,然后按一定的传播准则选取下一个数据点,调整该点处的法矢使得这两点的法矢保持一致,然后选取下一个新的点进行调整,直到完成所有点的法矢方向调整。由于前面已经建立了含有点云数据拓扑关系的 Riemannian 图,因此通过遍历 Riemannian 图的最小生成树(MST)可以完成点云法矢方向调整的过程。遍历的优先序列就是法矢调整的传播顺序。通过改变 Riemannian 图中边的权值,可以改变遍历最小生成树时的优先顺序,从而改变法矢调整的传播顺序。

利用双邻接矩阵表示 Riemannian 图不但可以节省存储空间,而且可以提高运算速度,因此是一种理想的存储方式。

然后估计出点云中各点处的法矢。点云数据中一点  $P$  的法矢  $N$  可以看作是点及其邻域  $nbhd(P)$  的最小二乘拟合平面的法矢。假设点  $O$  为点  $P$  的邻域  $nbhd(P)$  的重心,则:

$$O = \frac{1}{k} \sum_{i=1}^k P_{i1} \quad (48-1-9)$$

最小二乘拟合平面的法矢可以通过求解下列优化问题的最小值得到:

$$I = \sum_{i=1}^k \|(P_{i1} - O) \cdot N\|^2$$

该优化问题可以转化为求协方差矩阵的最小特征值问题:

可以看出,在法矢的调整过程中,数据点搜索(传播)的顺序是十分重要的。如果只根据数据点之间几何上距离最近作为法矢方向传播的优先准则,有可能引其错误的法矢方向调整。例如,在曲面变化比较大的地方(含有尖锐的棱边),有可能两点的距离很近,但它们的法矢方向却相反。如果按距离准则将两点的法矢方向调整为相同,那么显然是错误的。

因此,法矢方向的调整应当在法矢方向变化最小的两点之间传播。为 Riemannian 图中的每一条边增加一个权值  $\text{cost}$ , 设有两相邻数据点  $i, j$ , 边  $e(i, j)$  的权值  $\text{cost} = 1 - |\mathbf{n}_i \cdot \mathbf{n}_j| \geq 0$ , 两点之间的法矢方向变化越小,  $\text{cost}$  值越小, 遍历的优先级越高。该方法主要包括两个步骤: 一是建立 Riemannian 图的最小生成树, 二是遍历最小生成树。其实可以将上述两步合成一步完成。即直接利用最小生成树的实现算法达到对所有数据点的法矢进行调整的目的, 而不需要再遍历最小生成树。

该算法是在 Prim 算法的基础上实现的, 其基本思想是: 从 Riemannian 图的某一顶点  $x_i$  出发, 选择与它相邻的具有最小权值的边  $e(x_i, x_j)$ , 将顶点加入到生

成树的顶点集合  $U$  中, 比较这两点处的法矢, 如果  $\mathbf{n}_i \cdot \mathbf{n}_j < 0$ , 则将  $\mathbf{n}_j$  的方向取反。以后每一步从一个顶点在  $U$  中, 而另一个顶点不在  $U$  中的各边当中选择权值最小的边  $e(u, v)$ , 点  $u$  为集合  $U$  内的点, 即已经搜索过的点,  $v$  为集合外的点, 即需要调整的点。比较这两点处的法矢  $\mathbf{n}_u, \mathbf{n}_v$ , 如果  $\mathbf{n}_u \cdot \mathbf{n}_v < 0$ , 则将  $\mathbf{n}_v$  的方向取反, 并把  $v$  点加入到集合  $U$  中。如此继续下去, 直到网络中的所有顶点都加入到生成树的顶点集合  $U$  中为止。

可以看出, 点云数据中某一点处的法矢估计与该点的邻域选取有较大的关系, 与邻域点的个数  $k$  也有关系。当  $k$  值取的越小, 处理速度越快, 但算法的精度会降低; 当  $k$  值取的越大, 处理速度越慢, 但算法的精度可以提高。通常取  $k=25$ , 这样既可以做到效率和精度兼顾。

图 48-1-5 表示了点云数据的法矢调整过程, 图 48-1-5a 为一个球面的原始点云数据, 图 48-1-5b 为通过计算得到的初始点云数据各点处的法矢, 法矢的方向不一致, 有些地方法矢朝外, 另一些地方法矢朝内; 图 48-1-5c 为调整以后的法矢, 各点处的法矢一致朝外。

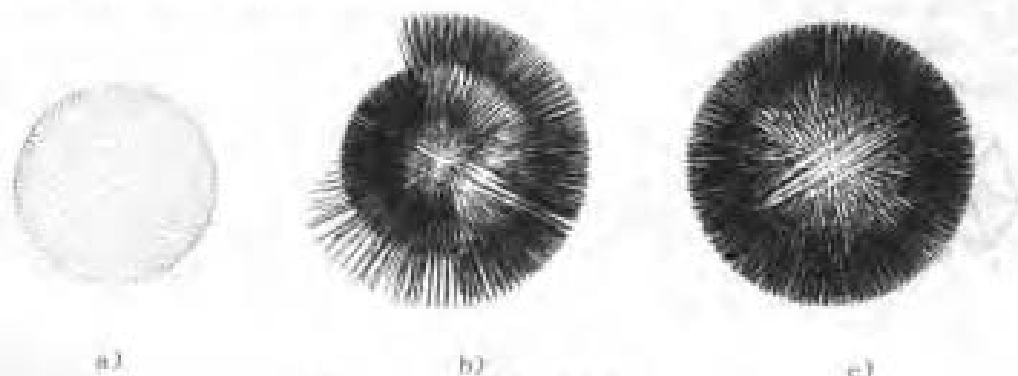


图 48-1-5 点云数据法矢调整

a) 点云数据 b) 初始法矢 c) 调整后的法矢

#### (4) 快速 Delaunay 三角剖分

目前散乱点云数据的三角网格曲面重建算法可以大体分为两类: 等值面方法和 Delaunay 三角剖分算法。

等值面方法将整个点云数据分割成容易处理的子点云, 对每一片子点云抽取等值面, 用它来近似表示原始子点云。这些等值面整合在一起就形成了原始点云的三角网格曲面。最典型的等值面方法是 Marching Cube 方法(简称 MC 方法), 它的基本原理是以分治法来计算三维数据场中每一个小立方体中的等值面。小立方体在数据场中以扫描线方式移动, 因此成为步进立方体。当立方体移动到某一个位置时, 计算该立方体 8 个顶点在数据场中的值, 然后根据线性插值的方法得到该立方体内的等值面, 将该等值面分割成若干三

角面片, 这些三角面片就近似表示该位置的体数据的形状。立方体不断移动, 就得到整个点云数据的几何形状。

由于等值面不一定通过所有的原始点云数据, 因此该方法得到的网格曲面只是物体的近似表示, 容易丢失一些细节部分的几何信息。

Delaunay 三角剖分算法是直接对散乱点云数据进行 Delaunay 三角剖分, 完成三角网格曲面的重建。一般采用区域增长算法, 即先从一个种子三角形开始, 不断将新的数据点加入到三角形区域内部, 按 Delaunay 准则进行三角剖分, 形成新的三角形区域。直到所有的数据点都加入进来, 完成整个点云数据的三角剖分过程。但是当散乱点的数目很大时, Delaunay 三角剖分计算量大, 算法执行速度慢。

采用改进的 Delaunay 三角剖分算法,实现快速三角网格剖分。在给出具体算法之前,为了方便理解,我们先给出一些定义。

**定义 1: 边界边** 在三角剖分过程中,位于已剖分区域与未剖分区域之间的三角形边称为边界边。如图 48-1-7a 所示边 1~6 为边界边。

**定义 2: 边界环** 在三角剖分过程中,由边界边按顺时针或逆时针方向依次首尾相连构成的环称为边界环。如图 48-1-7a 所示,边 1~6 按逆时针方向构成边界环。约定边界环按逆时针方向首尾相连。设数据点  $B$  的法矢为  $N_B$ , 向量  $V = \overrightarrow{AB} \times \overrightarrow{BC}$ , 令  $T = N_B \cdot V$ , 则  $T > 0$ 。

**定义 3: 可见点** 如图 48-1-7b 所示, 设点  $P$  位于边界边  $AB$  的右侧, 那么点  $P$  相对于边界边  $AB$  即为可见点。一条边界边可能有多个可见点, 如点  $Q$ , 点  $R$  都是边界边  $AB$  的可见点。某一点与某一点边的位置关系有三种, 点在边的左侧、右侧或者与边重合。判断点  $P$  与边  $AB$  的位置关系可以用如下的判定方法来实现。

如图所示, 设  $A$  点的法矢值为  $N_A$ , 向量  $V = \overrightarrow{AP} \times \overrightarrow{AB}$ , 令  $T = N_A \cdot V$ ,

若  $T > 0$ , 则, 点  $P$  在边  $AB$  的右侧;

若  $T < 0$ , 则, 点  $P$  在边  $AB$  的左侧;

若  $\det(A, B, P) = 0$ , 则点  $P$  与边  $AB$  重合;

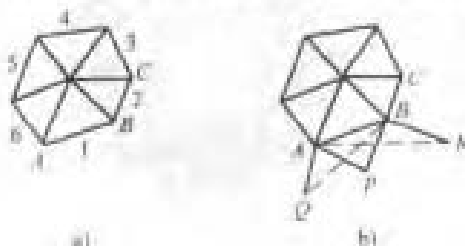


图 48-1-7 边界环与可见点

算法步骤如下:

第一步, 选取一个初始三角形, 一般由  $z$  坐标值最大的点及离它最近的两个点构成。选取三点  $p_{i-1}, p_i, p_{i+1}$  为逆时针排列的位于边界上的三角形顶点。

第二步, 将点  $p_{i-1}, p_{i+1}$  的邻接点  $p_{i-2}, p_{i+2} \in \text{Nbrhd}(p_{i-1}, p_{i+1})$  投影到点  $p_i$  处的切平面上, 得到投影点  $p_{i-2}', p_{i+2}'$ , 从投影点集中选取位于有向边  $p_{i-1} \rightarrow p_{i+1}$  右侧且满足下列条件的投影点  $p_{i-2}'$  作为判定条件:

$$\cos \angle p_{i-2}' p_i p_{i+1} = \min \left\{ \frac{(p_{i-2}' - p_i) \cdot (p_{i+1} - p_i)}{\|p_{i-2}' - p_i\| \|p_{i+1} - p_i\|} \right\} \quad (48-1-10)$$

该等式基于 Delaunay 三角化原则, 使  $\angle p_{i-2}' p_i p_{i+1}$  保持最大。

如果  $p_{i-2}'$  满足以下三种条件之一, 则其对应的原始点  $p_{i-2}$  可用来构成新三角形。

1)  $p_{i-2}$  为边界之外的点 如图 48-1-8a 所示, 数据点  $p_{i-2}, p_{i-1}, p_i$  构成新三角形  $\triangle p_{i-2} p_{i-1} p_i$ , 更新边界环, 在边界环中删去边界边  $\overrightarrow{p_{i-1} p_{i+1}}$ , 插入新边界边  $\overrightarrow{p_{i-2} p_{i+1}}, \overrightarrow{p_{i-2} p_{i-1}}$ 。

2)  $p_{i-2}$  在边界上, 且  $p_{i-2}$  为  $p_{i-1}$  的后继点 如图 48-1-8b 所示, 数据点  $p_{i-2}, p_{i-1}, p_i$  构成新三角形  $\triangle p_{i-2} p_{i-1} p_i$ , 更新边界环, 在边界环中删去边界边  $\overrightarrow{p_{i-1} p_{i+1}}, \overrightarrow{p_{i-1} p_{i-2}}$ , 插入新边界边  $\overrightarrow{p_{i-2} p_{i+1}}$ 。

3)  $p_{i-2}$  在边界上, 且  $p_{i-2}$  为  $p_{i-1}$  的前继点 如图 48-1-8c 所示, 数据点  $p_{i-2}, p_{i-1}, p_i$  构成新三角形  $\triangle p_{i-2} p_{i-1} p_i$ , 更新边界环, 在边界环中删去边界边  $\overrightarrow{p_{i-1} p_{i+1}}, \overrightarrow{p_{i-2} p_{i-1}}$ , 插入新边界边  $\overrightarrow{p_{i-2} p_{i+1}}$ 。

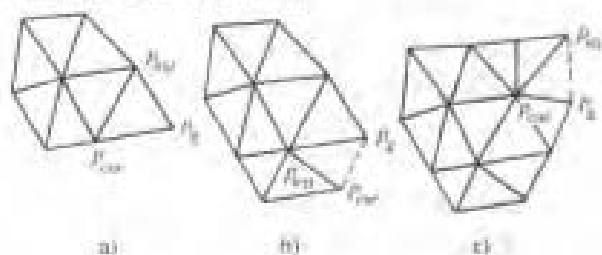


图 48-1-8 三角剖分

判断完毕, 更新网格边界环, 沿逆时针方向转向下一条边的判断, 如此迭代下去, 直到点云数据中所有点都进行了三角剖分或者边界环中边界边的个数为零为止。前一种情况为非封闭曲面的情形, 后一种情况为封闭曲面的情形。

图 48-1-9 为一个球面点云的三角剖分的结果, 图 48-1-9a 为原始散乱点云, 图 48-1-9b 为三角剖分后生成的三角网格曲面。

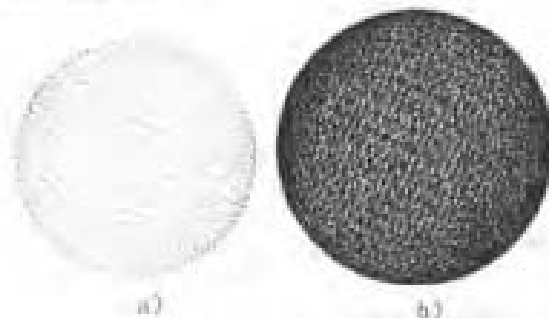


图 48-1-9 球面点云的三角剖分

### (5) 点云数据集的划分

通常一个物体表面是由多个单一曲面构成的, 包括二次曲面和自由曲面等, 因此在模型重构时, 需要将整个点云数据进行分割, 每一片点云用特定的曲面来拟合。在反求工程中, 通过测量获得的物体表面信息经过拟合、重采样等处理后得到的数据一般为散乱点云数据。基于空间散乱点云数据的物体模型重构, 一般利用手工的方法提取点云数据的边界, 然后利用这些边

界,将点云数据分割,对于每一片点云,手动选择一定的曲面进行拟合,重构零件的CAD模型。边界的提取和拟合曲面的选取都是人工的方法,这样不但精度不高,而且效率低。目前点云数据的自动分割主要集中在规则点云数据的研究,基于规则点云数据的自动分割其应用范围具有明显的局限性。也有一些学者进行了关于散乱点云数据的自动分割的研究,目前主要有两种点云数据的分割方法:基于边界的方法和基于面的方法。

如图48.1-10所示,物体表面的边界点可以分为三类:①位置不连续的点(或称为 $D^0$ 边界);②切矢(或法矢)不连续的点(或称为 $D^1$ 边界);③曲率不连续的点(或称为 $D^2$ 边界、光滑边界)。

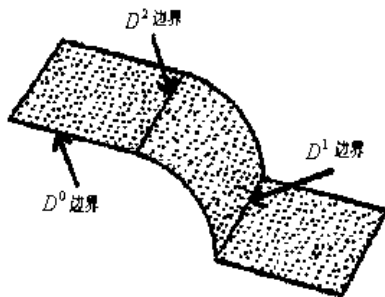


图 48.1-10 边界类型

基于边界的方法是先寻找曲面之间的边界点。在找到了边界点之后,再将这些边界点拟合出各分割点云之间的边界线,利用这些边界线及其被包围的分片点云最终拟合出曲面。基于边界的点云数据分割方法原理简单可行,可以利用人工交互的方法实现。但同时也存在以下问题:①敏感数据,特别是激光扫描得到的数据,常常恰好在清晰边界处不可靠;②可用于数据分割点的数目少,仅限于采用的边界点的范围内,这就意味着大量其他点的信息不可以用来辅助生成可靠的面片;③寻找光滑边界点,即切矢连续或更高阶连续的点,十分困难,这是由于存在噪点和测量误差引起的。另一方面,如果首先对数据进行光顺处理以减少误差,那么可能使计算出的点的曲率和法矢发生变化,而且,过滤噪点可能会移动特征的位置。

基于面的方法采用相反的过程来生成曲面,它从种子点开始,假设种子点及其邻域属于某一类型的曲面,在种子点周围逐渐加入属于这一曲面的邻接点,邻接点不断增长,直到没有属于当前区域的相邻的点,这些点构成了具有相同属性的点云连通区域,作为一片需要重构的曲面的原始点云数据。可以在零件表面的不同区域选取种子点,并行处理,生成不同的点云,然后通过合并、延伸、相交等操作来得到各片点云之间的边界。

基于面的技术具有以下优点:①它使用了更多的点,最大限度的利用了所有可以得到的数据。②可以直接确定哪些点属于哪些曲面。然而采用基于边界的方法,即使我们已经找到了一系列边界点,也不能完全确定给定的点属于哪片曲面。③直接提供了点云数据的最佳拟合曲面。

但基于面的方法也存在一些缺点:①很难选定最佳的种子点;②必须根据属于当前区域的点小心地更新假设,如果加入了坏点,那么会破坏对当前曲面属性的估计。③无法表示出一张复杂的自由曲面,而是可能会生成很多的细小平面或二次曲面,得不到预期的结果。

因此,一些学者提出将基于边界的分割技术与基于区域的分割技术结合的方法。利用基于边界与基于区域的分割技术相结合的方法对阵列点云进行分割:首先利用基于边界的分割技术,计算每一点处的法矢,利用 Kohonen 神经网络检测边界点;然后采用基于区域的分割技术,计算每一点处的平均曲率和高斯曲率,生成 K-H 标记图,输入 BP 网络,输出边界类型。最后结合基于边界的分割和基于区域的分割技术,得到最终的分割结果。

另一种方法是先将点云数据进行三角剖分。对每一个网格三角形的顶点,利用其邻域计算切平面,并根据欧拉公式计算任意方向的方向曲率,然后计算三角形任一边的垂直方向曲率及左右两顶点的方向曲率,通过比较确定该边是否为边界及哪一类边界。提出了“边界区域”(Border Region)概念,由特征边界附近的三角面片构成边界区域,利用边界区域将三角网格分成独立的网格区域。利用上述网格区域拟合特征曲面,最终实际的边界由特征曲面的布尔运算得到。

由于基于边界和区域相结合的分割方法结合了两种方法的优点,在对曲面进行分割的同时确定曲面的具体类型。具体方法是计算点云数据的高斯曲率和平均曲率,利用曲率极值作为边界点特征,进一步将边界点连接成边界曲线,利用边界曲线将点云数据进行自动分割。然后对每一片点云根据其曲率值分析该片点云进行几何形态分析,确定曲面的具体类型与相关几何形态参数和位置参数,为曲面重构作基础。

曲面的法矢和曲率是曲面的基本特性,也是曲面特征识别的重要依据之一,在点云数据特征提取与自动分割中,一般都要利用法矢与曲率等几何特征。曲面的一般形式可以表示为: $f(x, y, z) = 0$ 。假设曲面是光滑连续的,并且在每一点处都存在法矢。在曲面上任取一点,假设坐标系的 $z$ 轴方向与该点处的法矢方向接近,那么在该点处曲面可以用如下等式表示: $z = f(x, y)$ 。一般情况下,假定该点处为坐标原点,用 $u, v$ 重新



参数化。利用 Maclaurin 展开式得到:

$$f(u, v) = f_u(0, 0)u + f_v(0, 0)v + \frac{f_{uu}(0, 0)}{2}u^2 + f_{uv}(0, 0)uv + \frac{f_{vv}(0, 0)}{2}v^2 + \dots \quad (48.1-11)$$

将曲面定义为矢量形式:

$$S(u, v) = [uvf(u, v)]^T \quad (48.1-12)$$

为了简化公式表示, 记  $f_u = f_u(0, 0)$ ,  $f_v = f_v(0, 0)$ , 以此类推。

曲面在 origin 处的两个偏导数分别为:

$$S_u = [1 \ 0 \ f_u]^T, S_v = [0 \ 1 \ f_v]^T$$

二阶偏导数分别为:

$$S_{uu} = [0 \ 0 \ f_{uu}]^T,$$

$$S_{uv} = [0 \ 0 \ f_{uv}]^T,$$

$$S_{vv} = [0 \ 0 \ f_{vv}]^T$$

则 origin 处的单位法矢就可以表示为:

$$n = \frac{S_u \times S_v}{|S_u \times S_v|} = \frac{[-f_u \ -f_v \ 1]^T}{\sqrt{1 + f_u^2 + f_v^2}} \quad (48.1-13)$$

根据曲面第一、第二基本形式, 得:

$$E = S_u^T \cdot S_u = 1 + f_u^2,$$

$$F = S_u \cdot S_v = f_u f_v,$$

$$G = S_v^T \cdot S_v = 1 + f_v^2$$

$$L = S_{uu} \cdot n = \frac{f_{uu}}{\sqrt{1 + f_u^2 + f_v^2}},$$

$$M = S_{uv} \cdot n = \frac{f_{uv}}{\sqrt{1 + f_u^2 + f_v^2}},$$

$$N = S_{vv} \cdot n = \frac{f_{vv}}{\sqrt{1 + f_u^2 + f_v^2}}$$

$$\text{令: } g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix};$$

$$b = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} L & M \\ M & N \end{bmatrix}$$

则曲面在该点处沿任意方向  $u\delta = [du \ dv]^T$  的法曲率  $k_n$  可以表示如下:

$$k_n = \frac{u^T b u}{u^T g u} \quad (48.1-14)$$

利用公式(1-14)可以得出曲率极值即主曲率满足如下等式:

$$(b - k_n g)u\delta = 0 \quad (48.1-15)$$

公式展开为:

$$\begin{cases} (L - k_n E)du + (M - k_n F)dv = 0 \\ (M - k_n F)du + (N - k_n G)dv = 0 \end{cases} \quad (48.1-16)$$

消去  $du, dv$ , 得:

$$\begin{aligned} (EG - F^2)k_n^2 - (EN + GL - 2FM)k_n + \\ (LN - M^2) = 0 \end{aligned} \quad (48.1-17)$$

方程(1-17)的两个根即为两个主曲率  $k_{n1}, k_{n2}$ 。

将这两个主曲率  $k_{n1}, k_{n2}$  分别代入等式(1-16)可以求出对应的两个主方向  $u_1, u_2$ 。另外可以得到该点处的高斯曲率和平均曲率分别为:

$$\begin{aligned} K &= k_{n1} \cdot k_{n2} = \frac{LN - M^2}{EG - F^2} \\ &= \frac{f_{uu}f_{vv} - f_{uv}^2}{(1 + f_u^2 + f_v^2)^2} \quad (48.1-18) \\ H &= \frac{k_{n1} + k_{n2}}{2} = \frac{EN - 2FM + GL}{2(EG - F^2)} \\ &= \frac{(1 + f_v^2)f_{uu} + (1 + f_u^2)f_{vv} - 2f_u f_v f_{uv}}{2(1 + f_u^2 + f_v^2)^{3/2}} \end{aligned} \quad (48.1-19)$$

利用曲面的法矢和曲率这些几何性质, 我们就可以进行点云数据的特征提取与自动分割。

点云数据根据数据点之间的拓扑关系可以分为三类: 阵列点云、线状点云和散乱点云。阵列点云中的数据点呈规则的阵列形式排列, 具有“行×列”的特点, 激光扫描仪一次测量得到的数据就是一幅阵列点云。按线组织的数据可以看成是部分散乱的数据, 可认为是第三类数据的特例。利用阵列点云和散乱点云重构曲面的方法存在较大的差别。阵列点云数据的几何特征提取可以利用点云多视图拼合时给出的基于图像的方法实现, 这里我们主要研究散乱点云数据的几何特征提取与自动分割。

在求解散乱点云数据中某一点处的几何特征时, 一般利用二次曲面来逼近该点及其邻域的散乱点集, 然后通过计算该二次曲面的法矢和曲率来得到点云的法矢和曲率。

由于前面已经建立了散乱点云数据的三角网格曲面模型, 因此可以在该三角网格曲面的基础上计算曲率。如图 48.1-11a 所示, 假设在三角网格中有一点  $p$ , 所有以该点作为顶点的三角面片的集合为  $\Delta$ , 与点  $p$  相邻的点的集合为  $\{q_i\}, i=1, 2, \dots, 5$ 。设点  $p$  与点集  $\{q_i\}$  以及  $\{q_i\}$  的相邻点组成的点集为  $p_i$ , 我们利用该点集拟合最小二次曲面求解点云的曲率。

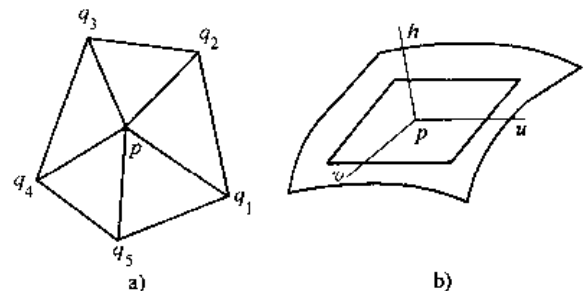


图 48.1-11 局部二次曲面拟合

首先, 在以  $p$  为原点建立局部坐标系, 如图 48.1-

11b所示,坐标轴为 $(u, v, h)$ ,以点 $p$ 的法矢方向作为其中一个坐标轴 $h$ 的方向,其余两个坐标轴 $u, v$ 在点 $p$ 的切平面中选取。结合公式(1-11),选取前面5项作为二次曲面的表示,则拟合最小二次曲面在局部坐标系中的表示为:

$$h = au + bv + \frac{c}{2}u^2 + duv + \frac{e}{2}v^2 \quad (48.1-20)$$

其中,

$$a = f_u, b = f_v, c = f_{uu}, d = f_{uv}, e = f_{vv}$$

设点集 $P, i=1, 2, \dots, n$ 中任意一点在局部坐标系中的对应参数为 $(u_i, v_i, h_i)$ ,建立方程组:

$$AX = B$$

其中,

$$A = \begin{bmatrix} u_1 & v_1 & \frac{u_1^2}{2} & u_1v_1 & \frac{v_1^2}{2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n & v_n & \frac{u_n^2}{2} & u_nv_n & \frac{v_n^2}{2} \end{bmatrix},$$

$$X = \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}, B = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$

求解方程组得:  $X = [A^T A]^{-1} A^T B$

利用上一节公式(1-13),可以得出该点处的法矢:

$$n = \frac{[-a \ -b \ 1]^T}{\sqrt{1+a^2+b^2}}$$

曲面第一、第二基本量为

$$E = 1 + a^2, F = ab, G = 1 + b^2,$$

$$L = \frac{c}{\sqrt{1+a^2+b^2}},$$

$$M = \frac{d}{\sqrt{1+a^2+b^2}},$$

$$N = \frac{e}{\sqrt{1+a^2+b^2}}$$

两个主曲率满足如下等式:

$$(1+a^2+b^2)k_n^2 - \left( \frac{(1+a^2)e + (1+b^2)c - 2abd}{\sqrt{1+a^2+b^2}} \right) k_n + \frac{ce - d^2}{1+a^2+b^2} = 0$$

求解该方程,可以得到两个主曲率,并将结果代入如下公式,可以进一步得出两个主曲率对应的主方向:

$$\begin{cases} (L - k_n E)du + (M - k_n F)dv = 0 \\ (M - k_n F)du + (N - k_n G)dv = 0 \end{cases}$$

利用公式(1-18)、(1-19)可以求出高斯曲率和平均曲率:

$$\text{高斯曲率: } K = \frac{ce - d^2}{(1+a^2+b^2)^2}$$

$$\text{平均曲率: } H = \frac{(1+a^2)e + (1+b^2)c - 2abd}{2(1+a^2+b^2)^{3/2}}$$

边界点可以认为是曲率极值点。同规则点云边界提取的方法类似,在提取边界点时,可以首先设定一个阈值 $\gamma$ ,如果曲面上某一点处的任一个方向的主曲率绝对值大于阈值,则作为候选边界点。这样就可以得到一系列候选边界点,实际边界点就在这些候选边界点内产生。

在这里,我们引入边界边的概念。在进行边界提取时,不是孤立地提取单个的边界点,而是一对边界点即边界边的形式提取边界。如图48.1-12所示,在两个候选边界点之间存在一条三角形边 $e$ ,设该边的边曲率为 $k_e$ 。与边 $e$ 对应的两个顶点 $v_3, v_4$ 的切平面上垂直于该边的方向分别记为 $t_3, t_4$ 。由于每一点处的主曲率和主方向已经得到,利用欧拉公式可以计算出顶点 $v_3, v_4$ 处沿 $t_3, t_4$ 方向的方向曲率,记为 $k_{v_3}, k_{v_4}$ 。比较 $k_e, k_{v_3}, k_{v_4}$ 之间的大小,可以确定边 $e$ 是否为边界边,具体方法如下。

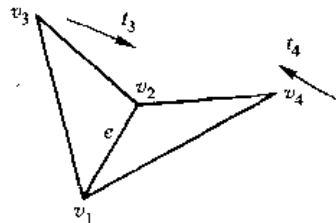


图 48.1-12 边界边

① 如果:  $|k_e| > |k_{v_3}|, |k_e| > |k_{v_4}|$ , 且  $|k_e - k_{v_3}| > t_1$  或  $|k_e - k_{v_4}| > t_1$ , 则边 $e$ 为 $D^1$ 边界边。

② 如果:  $\max(k_{v_3}, k_{v_4}) > k_e > \min(k_{v_3}, k_{v_4})$ , 且  $|k_{v_3} - k_{v_4}| > t_2$ , 则边 $e$ 为 $D^2$ 边界边。

式中,  $t_1, t_2$ ——设定的阈值。

通过上述方法,可以得到边界边的集合,集合内部的边界边是无序的。需要将这些提取出来的边界边连接成封闭的边界环。首先将边界边连接成若干独立的边界线,边界线的连接方法如下:

① 从边界边集合中选取一条边界边,作为新的边界线的初始线段,并将该边从集合中删除。

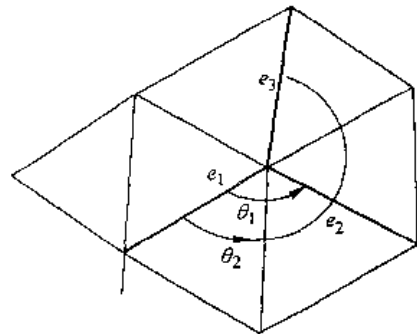


图 48.1-13 边界环连接

② 寻找与边界线具有共同端点的下一条边界边, 连接边界线与该边界边, 将新的边界边从边界集合中删除。

③ 重复过程②, 如果不能找到下一条边界边, 该边界线搜索结束, 形成一条边界线, 加入边界线集合, 跳到过程①。

④ 迭代下去, 直到边界边集合为空。

这样可以得到若干边界线的集合, 利用边界环连接算法可以将边界线连接成封闭的边界环。如图 48-1-13 所示, 首先获得一条边界线的最后一条边界边  $e_1$ , 然后寻找离它最近的另一条边界线  $e_2$  的端点,

将这两条边界线连接起来, 再从  $e_2$  开始, 寻找与它相邻的下一条边界线, 直到形成边界环。如果与  $e_1$  相邻的边界线有两条, 设为  $e_2, e_3$ , 计算  $e_1$  与  $e_2, e_3$  的夹角, 分别记为  $\theta_1, \theta_2$ , 如果  $\theta_1 < \theta_2$ , 那么将  $e_1$  与  $e_2$  相连。否则,  $e_1$  与  $e_3$  相连。

利用这些边界环就可以将整个点云数据自动分割。图 48-1-14 表示了点云数据的自动分割过程, 以鼠标点云数据作为实例。图 48-1-14a 为点云的最小主曲率分布, 图 48-1-14b 为点云的最大主曲率分布, 图 48-1-14c 显示了点云数据的边界环, 图 48-1-14d 显示了利用边界环分离出的其中一片点云。

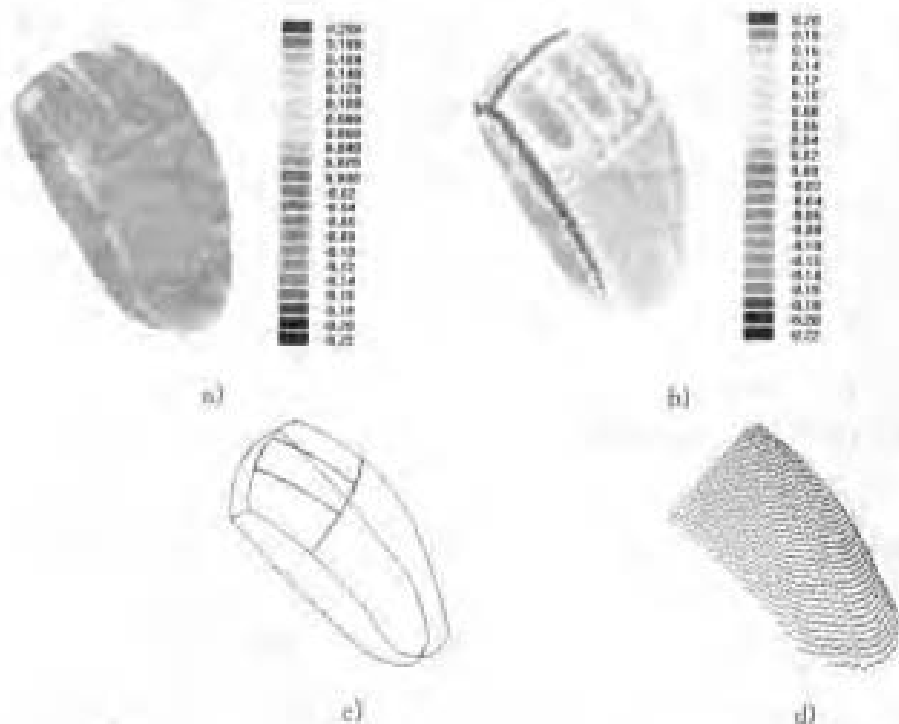


图 48-1-14 点云数据分割

### 2.3 曲面重构技术

点云数据经过分割处理后, 每一片分割后的点云需要用恰当类型的曲面来表示。基于边界的点云数据分割只能得到一些独立的点云片, 但每一片点云的具体类型不确定, 且目前大多通过人工交互的方法确定, 但这种方法容易发生偏差, 而且减低了效率。因此, 如果能自动准确地判定出点云的曲面类型, 则可以较好地完成点云的重建。

目前已有一些学者进行了曲面自动识别与分类方面的研究, 二次曲面根据系数可以确定曲面的具体类型, 但由于在曲面拟合时利用代数距离替代几何距离, 可能影响到曲面类型的正确判断。另一种曲面分类的

方法基于局部微分特性的估计。例如, 利用 Flynn 和 Jain 估计曲率值, 然后采用从统计学的角度对曲面进行分类。Foresti 提出一种基于神经树的方法识别曲面类型, 将曲面分为五类, 然后根据曲率特性对每一个数据点进行识别。然而, 单纯依靠基于曲率的分类方法不可避免地会受到噪声的影响。

曲面自动识别与分类是曲面重建的基础, 有关二次曲面和自由曲面重建的文献很多, 自由曲面的拟合一般利用 B 样条曲面或 NURBS 曲面实现。B 样条曲面可以看成是 NURBS 曲面的一个特例, 因此 NURBS 更具有通用性, 目前的商业 CAD 软件基本采用 NURBS 曲面作为自由曲面的表达方式。

根据几何复杂程度, 可以将曲面分成以下三类: 二

次曲面、扫掠曲面、自由曲面。二次曲面和扫掠曲面可以合称为规则曲面。规则曲面一般几何参数比较少,而自由曲面由于形态复杂,需要较多的几何参数才能精确地表示其几何形态。在反求工程的应用中,大多数产品表面为二次曲面,特别是平面、球面、圆柱面、锥面。采用局部几何形态分析的方法,直接利用点云内部的数据点的曲率特性,可以快速、自动地识别该点云是否为平面、球面或圆柱面。对于更一般的二次曲面,例如椭球面、抛物面、双曲面等,可利用全局几何形态的方法,对点云进行一般二次曲面拟合,根据曲面参数,判断该点云的曲面类型。曲面分类见图 48.1-15。

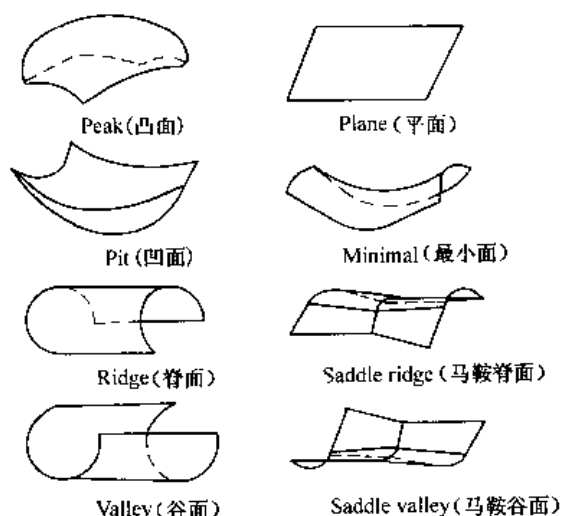


图 48.1-15 曲面分类

(1) 点云数据的几何形态分析

1) 局部几何形态分析 局部几何形态分析就是利用点云数据的法矢和曲率等几何信息,分析该点云所表示的曲面的局部几何形态,并确定曲面的具体类型,以便于点云数据的分割和曲面重构,国内外在这方面已有一定的研究。Besl 和 Jain 根据曲面的高斯曲率和平均曲率将曲面分成八类。具体如下:

- ①  $k_{gauss} > 0, k_{mean} < 0$  peak 曲面
- ②  $k_{gauss} > 0, k_{mean} > 0$  pit 曲面
- ③  $k_{gauss} = 0, k_{mean} < 0$  ridge 曲面
- ④  $k_{gauss} = 0, k_{mean} > 0$  valley 曲面
- ⑤  $k_{gauss} = 0, k_{mean} = 0$  plane 曲面
- ⑥  $k_{gauss} < 0, k_{mean} = 0$  minimal 曲面
- ⑦  $k_{gauss} < 0, k_{mean} < 0$  saddle ridge 曲面
- ⑧  $k_{gauss} < 0, k_{mean} > 0$  saddle valley 曲面

但这样的分类过于粗略,只能了解曲面的大致几何形态,不能直接用于曲面重构。若利用点云数据中每一点处的主曲率来分析曲面的类型,可以直接判断出某一片点云是否为平面、球面或圆柱面。

根据曲面的几何特性,平面上任意一点处的两个主曲率都为零。圆柱面上任意一点处的最小主曲率为零,最大主曲率为一个常数;球面上任意一点处的两个主曲率相同且为常数。考虑到被测物体的精度和存在一定的测量误差,对平面的识别,我们利用假设检验的方法来进行判断。

假设点云数据中任意点  $i$  的最大、最小主曲率分别为  $k_{min,i}, k_{max,i}, i=1, \dots, n, n$  为该点云中数据点的个数。建立两个检验假设:

- ①  $H_0: \bar{k}_{min} = 0, H_1: \bar{k}_{min} \neq 0$
- ②  $H_0: \bar{k}_{max} = 0, H_1: \bar{k}_{max} \neq 0$

其中,  $\bar{k}_{min} = \frac{1}{n} \sum_{i=1}^n k_{min,i}, \bar{k}_{max} = \frac{1}{n} \sum_{i=1}^n k_{max,i}$

构造两个统计量:

$$T_1 = \frac{\bar{k}_{min}}{S_{n,min}^*} \sqrt{n}, T_2 = \frac{\bar{k}_{max}}{S_{n,max}^*} \sqrt{n}$$

其中,  $S_{n,min}^* = \frac{1}{n-1} \sum_{i=1}^n (k_{min,i} - \bar{k}_{min})^2, S_{n,max}^* = \frac{1}{n-1} \sum_{i=1}^n (k_{max,i} - \bar{k}_{max})^2$

当原始点云为平面时,这两个统计量  $T_1, T_2$  都服从自由度为  $n-1$  分布。于是,对于给定的  $\alpha$ ,我们可以求得  $T_1, T_2$  的值  $t_1, t_2$ ,

若  $|t_1| \geq t_{\alpha/2}(n-1)$  或  $|t_2| \geq t_{\alpha/2}(n-1)$ ,则拒绝点云为平面的假设;

若  $|t_1| < t_{\alpha/2}(n-1)$  且  $|t_2| < t_{\alpha/2}(n-1)$ ,则接受假设,即点云为平面。

当  $n > 30$  时,  $t$  分布可以近似看作标准正态分布。

若  $|t_1| < t_{\alpha/2}(n-1)$ , 但  $|t_2| > t_{\alpha/2}(n-1)$ , 则表明点云的最小平均主曲率为零,但最大平均主曲率不为零。进一步判定点云是否为圆柱面,如果  $|\bar{k}_{max}| > \delta_0$ , 且  $S_{n,max}^* < \epsilon_0$ , 则可以认为点云上每一点处的最大主曲率基本为一个定值,故可以认为点云是圆柱面。其中,  $\delta_0, \epsilon_0$  为设定的阈值。根据实践经验,通常给定  $\delta_0 = 0.005\text{mm}, \epsilon_0 = 0.01\text{mm}$ 。

若  $|t_1| > t_{\alpha/2}(n-1)$  且  $|t_2| > t_{\alpha/2}(n-1)$ , 则表明点云的最小平均主曲率和最大平均主曲率均不为零。进一步判定点云是否为球面,如果  $|\bar{k}_{min} - \bar{k}_{max}| < \delta_1$ , 而且  $S_{n,min}^* < \epsilon_1, S_{n,max}^* < \epsilon_2$ , 则可以认为点云上每一点的最小平均主曲率和最大平均主曲率基本为定值且二者相等,故可以认为点云是球面。通常给定  $\delta_1 = 0.005\text{mm}, \epsilon_1 = \epsilon_2 = 0.01\text{mm}$ 。

图 48.1-16 为平面点云数据的局部几何形态分析的实例,求解该点云数据的最小、最大平均主曲率的均值与标准差为:  $\bar{k}_{min} = -0.0014, S_{n,min}^* = 0.0029, \bar{k}_{max} = 0.0014, S_{n,max}^* = 0.0028$ 。根据判断准则可以确定该曲

面为平面。

图 48-1-17 为圆柱面点云数据的局部几何形态分析的实例,求解该点云数据的最小、最大平均主曲率的均值与标准差为: $\bar{k}_{\min} = 5.89e-6, S_{\bar{k}_{\min}}^2 = 0.0029; \bar{k}_{\max} = 0.0089, S_{\bar{k}_{\max}}^2 = 0.0051$ 。根据判断准则可以确定该曲面为圆柱面。

图 48-1-18 为球面点云数据的局部几何形态分析的实例,求解该点云数据的最小、最大平均主曲率的均值与标准差为: $\bar{k}_{\min} = 0.0785, S_{\bar{k}_{\min}}^2 = 0.0035; \bar{k}_{\max} = 0.0824, S_{\bar{k}_{\max}}^2 = 0.0024$ 。根据判断准则可以确定该曲面为球面。

2) 全局几何形态分析 直接利用点云数据的主

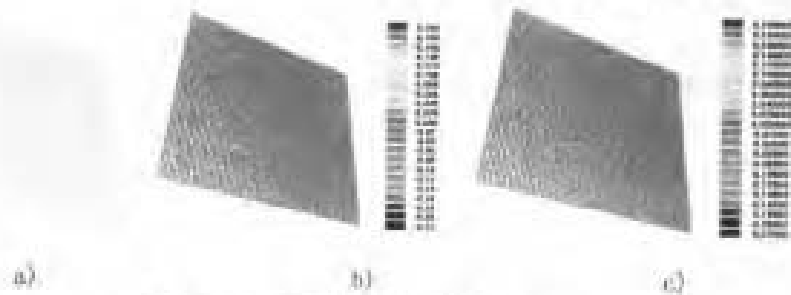


图 48-1-16 平面局部几何形态分析  
a) 点云数据 b) 最小主曲率 c) 最大主曲率



图 48-1-17 圆柱面点云数据  
a) 点云数据 b) 最小主曲率 c) 最大主曲率



图 48-1-18 球面点云数据  
a) 点云数据 b) 最小主曲率 c) 最大主曲率

曲率,我们可以识别出一部分二次曲面,但对于识别更一般的二次曲面,就比较困难。因此,采用全局几何形态分析的方法,识别一般的二次曲面。所谓全局几何形态分析,就是对分割后的每一片点云,利用一张一般二次曲面进行拟合,分析拟合误差,如果拟合误差在一定的范围内,那么可以认为该曲面为二次曲面,然后根据二次曲面的系数,判断该曲面的具体类型,如果拟合误差超出一定的范围,那么将该曲面作为自由曲面来重构。

(2) 一般二次曲面拟合

假设存在一片点云  $P = \{p_i\}, i = 1, \dots, n$ , 利用一张一般二次曲面进行拟合,一般二次曲面可以表示为:

$$F(x, y, z) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + 2a_{41}x + 2a_{42}y + 2a_{43}z + a_4 = 0 \quad (48-1-21)$$

式中,  $a_{ij} (i, j = 1, 2, 3, 4)$  是实数。

利用最小二乘法确定其中的 10 个系数,将拟合误差平方和作为优化目标函数,假定点云数据中点的个

数为  $n$ , 那么目标函数为:

$$e = \min \sum_{i=1}^n d_i^2 \quad (48.1-22)$$

这里,  $d_i$  表示点云数据中一点到拟合二次曲面的距离。假设点云数据中有一点  $p(x_0, y_0, z_0)$ , 该点处的法矢为  $n(n_x, n_y, n_z)$ , 那么点到二次曲面的距离就可以用该点沿法矢方向  $n$  到曲面的距离来表示。具体方法如下:

引入如下记号:

$$\begin{cases} F_1(x, y, z) = a_{11}x + a_{12}y + a_{13}z + a_{14} = \frac{1}{2} \frac{\partial F}{\partial x} \\ F_2(x, y, z) = a_{21}x + a_{22}y + a_{23}z + a_{24} = \frac{1}{2} \frac{\partial F}{\partial y} \\ F_3(x, y, z) = a_{31}x + a_{32}y + a_{33}z + a_{34} = \frac{1}{2} \frac{\partial F}{\partial z} \\ F_4(x, y, z) = a_{14}x + a_{24}y + a_{34}z + a_{44} \end{cases} \quad (48.1-23)$$

则,

$$F(x, y, z) = xF_1(x, y, z) + yF_2(x, y, z) + zF_3(x, y, z) + F_4(x, y, z)$$

$$\begin{cases} \Phi_1(x, y, z) = a_{11}x + a_{12}y + a_{13}z \\ \Phi_2(x, y, z) = a_{21}x + a_{22}y + a_{23}z \\ \Phi_3(x, y, z) = a_{31}x + a_{32}y + a_{33}z \\ \Phi(x, y, z) = x\Phi_1(x, y, z) + y\Phi_2(x, y, z) + z\Phi_3(x, y, z) \end{cases} \quad (48.1-24)$$

过点  $p$  且方向为  $n$  的直线的参数方程为:

$$\begin{cases} x = x_0 + n_x t \\ y = y_0 + n_y t \\ z = z_0 + n_z t \end{cases} \quad (48.1-25)$$

将方程式(1-25)代入方程式(1-21)得:

$$at^2 + bt + c = 0 \quad (48.1-26)$$

其中

$$\begin{cases} a = \Phi(n_x, n_y, n_z) \\ b = 2(n_x F_1(x_0, y_0, z_0) + n_y F_2(x_0, y_0, z_0) + n_z F_3(x_0, y_0, z_0)) \\ c = F(x_0, y_0, z_0) \end{cases} \quad (48.1-27)$$

求解方程式(1-26), 取其中绝对值最小的根作为点到曲面的距离, 如果没有实数根, 则表示直线与曲面没有交点, 可以不考虑该点。

方程式(1-22)是非线性最小二乘问题, 利用迭代优化方法求解。设定一个阈值  $\delta$ , 如果拟合误差  $e$  超过设定的阈值  $\delta$ , 那么认为该片点云不是单一的二次曲面, 应该利用自由曲面来重构, 如果拟合误差  $e$  小于阈值  $\delta$ , 那么认为该片点云是一张二次曲面, 下一步就是

要确定该二次曲面的具体类型。

1) 曲面类型的确定 通过求解该优化模型得到二次曲面的 10 个系数。下一步就是确定该二次曲面的具体类型, 我们利用不变量来确定二次曲面的类型。

定义 1.1: 不变量 由曲面方程的系数给出的函数, 如果在经过任意一个直角坐标变换后, 它的函数值不变, 就称这个函数是该曲面的正交不变量, 简称不变量。

不变量反映了曲面的内在几何特性, 它不随曲面在坐标系中的位置的改变而改变。设二次曲面的方程为(1-21), 记

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

$$\bar{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

设矩阵  $\bar{A}$  的特征值为  $\lambda_1, \lambda_2, \lambda_3$ , 引入四个不变量:

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3$$

$$I_2 = \lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_3 \lambda_1$$

$$I_3 = \lambda_1 \lambda_2 \lambda_3$$

$$I_4 = \det A$$

除了以上不变量, 我们还引入半不变量的概念。

定义 1.2: 半不变量 在保持原点不动的直角坐标变换下的不变量, 称为半不变量。

引入两个半不变量,

$$K_1 = \begin{vmatrix} a_{11} & a_{14} \\ a_{14} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{24} \\ a_{24} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{33} & a_{34} \\ a_{34} & a_{44} \end{vmatrix}$$

$$K_2 = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{12} & a_{22} & a_{24} \\ a_{14} & a_{24} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} & a_{14} \\ a_{13} & a_{33} & a_{34} \\ a_{14} & a_{34} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{23} & a_{33} & a_{34} \\ a_{24} & a_{34} & a_{44} \end{vmatrix}$$

利用这些不变量和半不变量, 我们将二次曲面简化成五大类, 并进一步分析二次曲面的类型。二次曲面可以分成 17 种曲面, 其中有一些是虚曲面, 而在反求工程中, 通过点云重构的曲面一定是实际存在的曲面。因此, 我们只列出在点云重构时可能出现的二次曲面情况, 其余曲面可以参考有关文献。

$$\textcircled{1} \text{ 当 } I_3 \neq 0 \text{ 时, 简化成: } \lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + \frac{I_4}{I_3} = 0$$

a) 若  $I_4 \neq 0$ ,

i 如果  $\lambda_1, \lambda_2, \lambda_3$  同号, 但与  $\frac{I_4}{I_3}$  异号, 曲面为椭球

面。

ii 如果  $\lambda_1, \lambda_2$  同号, 但与  $\lambda_3, \frac{I_4}{I_3}$  异号, 曲面为单叶双曲面。

iii 如果  $\lambda_1, \lambda_2, \frac{I_4}{I_3}$  同号, 但与  $\lambda_3$  异号, 曲面为双叶双曲面。

b) 若  $I_4 = 0$ , 如果  $\lambda_1, \lambda_2$  同号, 但与  $\lambda_3$  异号, 曲面为一阶锥面。

② 当  $I_3 = 0, I_4 \neq 0$  时, 简化成:  $\lambda_1 x^2 + \lambda_2 y^2 \pm 2\sqrt{-\frac{I_4}{I_3}}z = 0$ ;

a) 若  $I_4 < 0$ , 曲面为椭圆抛物面;

b) 若  $I_4 > 0$  则曲面为双曲抛物面, 即马鞍面。

③ 当  $I_3 = I_4 = 0, I_2 \neq 0$  时, 简化成:  $\lambda_1 x^2 + \lambda_2 y^2 + \frac{K_2}{I_2}z = 0$ ;

a) 若  $K_2 \neq 0$ ,

i 如果  $\lambda_1, \lambda_2$  都和  $K_2$  异号, 曲面为椭圆柱面。

ii 如果  $\lambda_1, \lambda_2$  异号, 曲面为双曲柱面。

b) 若  $K_2 = 0$ , 如果  $\lambda_1, \lambda_2$  异号, 曲面为两个相交平面。

④ 当  $I_2 = I_3 = I_4 = 0, K_2 \neq 0$  时, 简化成:  $\lambda_1 x^2 \pm 2\sqrt{-\frac{K_2}{I_1}}y = 0$ ; 曲面为抛物柱面。

⑤ 当  $I_2 = I_3 = I_4 = K_2 = 0$  时, 简化成:  $\lambda_1 x^2 + \frac{K_1}{I_1}z = 0$ ; 曲面为平面。

2) 曲面的中心与主方向 一般二次曲面通过化简之后, 可以较清楚地知道该二次曲面为何种类型的曲面。但有时还需要进一步确定其在空间中所处的位置, 如对称中心、对称轴等等。这些都是我们需要关注的对象, 下面给出二次曲面的中心和主方向的求解方法。

定理 1.1: 点  $C(x_0, y_0, z_0)$  是二次曲面  $S$  中心的充要条件是:

$$\begin{cases} F_1(x_0, y_0, z_0) = 0 \\ F_2(x_0, y_0, z_0) = 0 \\ F_3(x_0, y_0, z_0) = 0 \end{cases} \quad (48.1-28)$$

方程组展开为:

$$\begin{cases} a_{11}x_0 + a_{12}y_0 + a_{13}z_0 + a_{14} = 0 \\ a_{12}x_0 + a_{22}y_0 + a_{23}z_0 + a_{24} = 0 \\ a_{13}x_0 + a_{23}y_0 + a_{33}z_0 + a_{34} = 0 \end{cases}$$

$$\text{令 } \bar{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, X = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, U = \begin{bmatrix} -a_{14} \\ -a_{24} \\ -a_{34} \end{bmatrix}$$

方程组可以写成:

$$\bar{A}X = U \quad (48.1-29)$$

若方程组无解, 则曲面没有中心, 称此曲面为无心曲面。如果方程组有唯一解, 那么  $S$  有唯一中心, 这种二次曲面称为中心曲面。

若方程组存在唯一解, 求解该方程组可以得到曲面的中心:

$$X = [\bar{A}^T \bar{A}]^{-1} \bar{A}^T U \quad (48.1-30)$$

球面、椭球面、单叶、双叶双曲面为中心曲面; 抛物面为无心曲面。下面给出主方向的定义:

定义 1.3: 二次曲面  $S$  的奇向及  $S$  的对称面的法向称为  $S$  的主方向。

二次曲面  $S$  的主方向为矩阵  $\bar{A}$  的特征方向, 且不同特征根对应的主方向互相垂直。因此求解二次曲面  $S$  的主方向的方法如下:

1) 求出矩阵  $\bar{A}$  的特征根;

2) 求出特征根对应的主方向。

(3) 点云数据曲面重建

通过对点云数据进行局部和全局几何形态分析, 我们可以确定点云的具体曲面类型, 并且可以初步确定该曲面的几何形态参数和位置参数。利用这些信息, 可以较准确地重构点云数据。点云数据重构时, 一般将点云分为两类: 二次曲面和自由曲面。对于二次曲面, 可以直接采用最小二乘法来拟合。平面和球面的重构比较简单, 可以直接利用线性最小二乘算法求解。

1) 平面重构 设存在点云  $P = \{p_i\}, i = 0, 1, \dots, n$ , 拟合平面方程为:

$$ax + by + cz + d = 0 \quad (48.1-31)$$

设点集  $P$  中任意一点的坐标值为  $(x_i, y_i, z_i)$ , 建立方程组:

$$AX = 0$$

其中

$$A = \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{bmatrix}, X = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

矩阵  $(A^T A)$  的绝对值最小的特征值所对应的特征向量即为平面系数  $(a, b, c, d)$  的最小二乘解。

2) 球面重构 设拟合球面方程为:

$$(x + a)^2 + (y + b)^2 + (z + c)^2 = r^2 \quad (48.1-32)$$

展开为:

$$x^2 + y^2 + z^2 + 2ax + 2by + 2cz + (a^2 + b^2 + c^2 - r^2) = 0 \quad (48.1-33)$$

令:  $a^2 + b^2 + c^2 - r^2 = d$

代入方程式(1-22)得:

$$x^2 + y^2 + z^2 + 2ax + 2by + 2cz + d = 0$$

建立方程组:

$$AX = B$$

$$A = \begin{bmatrix} 2x_1 & 2y_1 & 2z_1 & 1 \\ 2x_2 & 2y_2 & 2z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 2x_n & 2y_n & 2z_n & 1 \end{bmatrix}, X = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$B = \begin{bmatrix} -(x_1^2 + y_1^2 + z_1^2) \\ -(x_2^2 + y_2^2 + z_2^2) \\ \vdots \\ -(x_n^2 + y_n^2 + z_n^2) \end{bmatrix}$$

求解该方程组,得:

$$X = [A^T A]^{-1} A^T B \quad (48-1-34)$$

从球面方程可知,球心坐标为 $(-a, -b, -c)$ ,球面半径为 $r, r = \sqrt{a^2 + b^2 + c^2 - d}$ 。

圆锥面、圆柱面及其他二次曲面的拟合则需要利用非线性最小二乘法求解。根据曲面方程,建立目标函数。假设曲面方程为 $S(x_1, \dots, x_n)$ ,点云数据为 $P = \{p_i | i = 1, \dots, n\}$ ,计算点云到拟合曲面的距离的残差平方和:

$$K = \sum_{i=1}^n d(S, p_i)^2 \quad (48-1-35)$$

利用非线性最小二乘法求解,可以得到曲面 $S$ 的参数。

3) 自由曲面重构 点云数据的自由曲面拟合,一般利用以下三种曲面类型拟合:①扫描曲面;②旋转曲面;③蒙皮曲面。针对不同的曲面采用不同的方法来拟合。国内外学者在这方面已经取得了较有成效的工作,特别是在B样条曲面或NURBS自由曲面重构方面。一种是利用离散点云重构旋转曲面和扫描曲面的方法。旋转曲面拟合的具体方法是利用迭代优化过程来降低测量点云与拟合曲面之间的平方根误差,得到优化后的旋转轴和最佳拟合截面曲线,利用旋转轴和截面曲线就可以建立旋转曲面的数学模型。扫描曲面拟合作为非线性最小平方拟合问题,通过迭代优化求解误差函数,得到最优的生成线、方向线和每一个测量点对应的参数值。另一种是利用NURBS曲面来拟合自由曲面,考虑NURBS在各阶边界约束条件(位置、切矢、曲率等)来拟合曲面。由于B样条曲面拟合具有操作方便、通用性强等特点,采用B样条曲面来拟合一般自由曲面,NURBS曲面可以看成是B样条曲面的进一步推广和延伸,NURBS曲面引入了权因子,因此使用范围更广,但同时也更加复杂,NURBS曲面拟合时,权因子的选择仍是一个需要深入研究的问题。B样条曲面拟合可以分为曲面插值与逼近两类。曲面插值要求重构的曲面严格通过数据点,当点云数据存在噪点时,生成的曲面光滑性很差,因此在实际应用中,较少采用插值的方法,一般采用逼近的方法进行曲面重

构。下面介绍B样条曲面重构自由曲面的方法。

双三次B样条曲面的逼近方法类似于B样条曲线的逼近,可以看成是曲线逼近的延伸。一张双三次B样条曲面可以用如下方程来表示:

$$S(u, v) = \sum_{i=1}^m \sum_{j=1}^n d_{ij} N_{i,3}(u) N_{j,3}(v) \quad u, v \in [0, 1] \quad (48-1-36)$$

式中  $d_{ij}$ ——双向控制顶点网格;  
 $u, v$ ——双向节点参数;

$N_{i,3}(u), N_{j,3}(v)$ —— $u, v$ 两个方向的B样条基函数。

B样条曲面的逼近可以看成:给定一片具有 $l+1$ 个数据点的点云 $p_i, i = 0, 1, \dots, l$ ,生成一张具有 $(m+1) \times (n+1)$ 个控制顶点的B样条曲面,使该B样条曲面与给定点云数据之间的距离平方和 $K$ 达到最小。用等式表示如下:

$$\min K = \sum_{i=0}^l [p_i - S(u, v)]^2 \quad (48-1-37)$$

利用非线性最小二乘的求解方法,可以得到该方程的解。求解时需要确定每一个数据点的初始双向参数,可以采用基平面投影法得到,将数据点投影到一张基平面,计算投影点的双向参数,然后以该参数作为数据点的参数,通过线性最小二乘法求解控制顶点。

定义控制顶点列向量 $d = [d_{11}, d_{12}, \dots, d_{1n}, d_{21}, d_{22}, \dots, d_{2n}, \dots, d_{m1}, d_{m2}, \dots, d_{mn}]^T$ ,同时定义对应的B样条基函数列向量 $T = [N_{i,3}(u) N_{j,3}(v)]^T, i = 0, \dots, m; j = 0, \dots, n$ ,可以得到如下方程:

$$T \cdot d = p \quad (48-1-38)$$

式中  $P = [p_1, p_2, \dots, p_n]^T$  为数据点列向量。解该方程可得:



图 48-1-19 点云数据自由重构



$$\mathbf{d} = (\mathbf{T}^T \cdot \mathbf{T})^{-1} \cdot \mathbf{T}^T \cdot \mathbf{p} \quad (48.1-39)$$

这样就可以得到控制顶点的值。由于数据点的初始参数不一定与曲面上对应最近点的参数相匹配,因此,需要通过调整数据点的参数,使式(1-37)的值达到最小,利用 Hoschek 算法可以得到新的数据点的参数,代入式(1-39)求解新的控制顶点,如此迭代下去,直到  $E < \epsilon$ , 得到最优的控制顶点。图 48.1-19 表示了点云的曲面重构整个过程,48.1-19a 为原始点云数据,图 48.1-19b 为点云的曲率分布,图 48.1-19c 为点云的边界线,图 48.1-19d 为曲面重建后的三维模型。该物体包括平面、椭圆柱面、自由曲面等特征。

## 2.4 实体构造技术

由曲面表面模型构造实体模型是反求工程技术中难点,目前的 CAD 软件几何都没有有效的办法加以解决。目前大多靠技术人员的经验和技巧,尤其对复杂曲面在实体重构中的自交和缝隙,在重构实体模型中,经常容易出现的问题有:

1) 缝隙 由于曲面拼合时很难保证曲面边界的统一和光滑过渡,以及曲面剪裁时带来的边界表示的特殊性,曲面之间可能存在很小的、不易发现的缝隙,不能达到实体的水密性要求。

2) 面片平移时的自相交 为形成零件内表面,经常需要对曲面平移,平移曲面时,在多曲面拼接的情况时极有可能出现曲面的自相交问题,而目前的 CAD 软件几乎都没有很好的解决办法。

3) 曲面的不可操作性 拟合出的各张曲面,仅仅只能控制其参数域范围内的形状,在进行延伸、圆角等操作时会出现形状畸变或自相交等问题。

4) STL 格式化 在将重构结果用于快速原型时,需要将组成零件表面的各曲面用 STL 格式进行三角面片化,相邻曲面进行三角面片化时易出现三角面片重叠、错边、缺失等问题。现有的 CAD 软件这方面处

理都不够理想,需要人工对处理结果进行修正。

5) 不同软件对曲面表示的不兼容问题 如我们所使用的 SURFACER 软件(IMAGEWARE 公司的产品)和各种 CAD 软件可能在以下几个方面出现问题:①NURBS 曲面表示方法②裁减曲面边界表示③曲面模型参数化表示。

## 3 反求工程的应用领域

### (1) 老产品复制

某些情况下,需要对某一产品进行原样复制,但又没有其图纸或 CAD 数据。用传统的方法对实物进行测绘再绘制工程图需要花费很长时间,尤其对含有复杂自由曲面的零件,对其精确测量几乎不可能,而采用反求工程技术处理这种情况非常适合。

### (2) 对已有产品的改进

产品在使用过程中由于磨损或不符合使用要求需改进的个别零件,特别如冲压件、塑料件与其他零件配合时,出现缝隙过大或表面不平,需要调整时,首先需获取配合零件的几何数据,进而在 CAD 系统中进行分析和改进。如汽车尾灯与车身的匹配的改善。

### (3) 新产品设计

在设计新产品时,往往需要参考已有的产品,它可以是老产品,也可以是市场上的同类产品。对于外观设计,有时会利用其他产品的某些外表曲面,或手工油泥模型。这些都需要由反求工程获取其 CAD 数据供新产品设计参考,应用反求工程进行产品设计的一般流程如图 48.1-20。

### (4) 产品验证

产品制造完成以后,用反求工程方法重构出该产品的 CAD 模型,与已有标准的 CAD 数据进行比较,分析误差。特别是在模具和快速成形等领域,工业界已用反求工程来定期地抽样检验产品,分析制造误差的规律,作为质量控制和分析产品缺陷的有力工具:

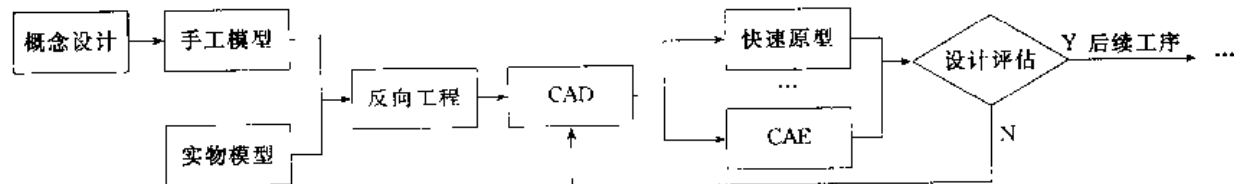


图 48.1-20 应用反求工程进行新产品设计的一般流程

### (5) 快速制造

根据反求方法测量的数据直接生成 STL 格式和 NC 代码,用快速原型技术制作样件,由样件通过快速制模翻成零件或由数控机床直接加工出零件。这样,能大大缩短零件制作周期,进而实现产品的快速制造。

反求工程技术提供了一种产品创新设计的新方

法,它不仅能实现传统技术难以实现的功能,如复杂油腻模型的曲面重构,而且可以大大缩短产品设计时间,加速新产品上市的速度。据资料报道和实例验证,应用反求工程技术后,产品的设计周期可以从几个月缩短为几周。反求工程也是支持敏捷制造、计算机集成制造、并行工程等的有效工具,是企业缩短产品开发周

期、降低设计生产成本、提高产品质量、增强产品的竞争力的关键技术之一。因而,这一技术已成为产品创新设计的强有力的支撑技术。

随着反求工程技术的不断发展,其应用的工业领域愈来愈广泛。除了在汽车、航空、家电等行业得到广泛应用外,在其他工程学科中,如医学领域,反求工程也得到了实际应用。例如:人工关节的设计和制造需要对骨骼周围的形状精确识别,然后生成数字化的 CAD 模型。目前主要使用 CT 技术,由 CT 图像获得骨骼结构的二维轮廓,然后重构出曲面模型供人工关节设计和制造。同样的方法已被用来分析化石中的原始人类和现代人的解剖学差异。反求工程还可以用来复制工艺品和修复已损坏的工艺品。电影和娱乐业还可以用

之帮助设计演员造型。在虚拟现实研究中,反求工程还可用来构造虚拟环境。通过数字化和 CAD 建模可以得到物体的数学表示,以创造一个更真实而细腻的虚拟世界。

#### 4 反求测量系统简介

有多种测量技术可以实现产品的数字化。常见的测量方法可分为接触式测量和非接触式测量。在医学界和一些工业领域,用医用 CT 和工业 CT 对实物进行断层扫描也是一种常用的方法。为了获取零件内部如内腔、孔等的的数据,还可采用 CGI 技术破坏式地切层扫描零件。这些测量技术的分类如图 48.1-21。

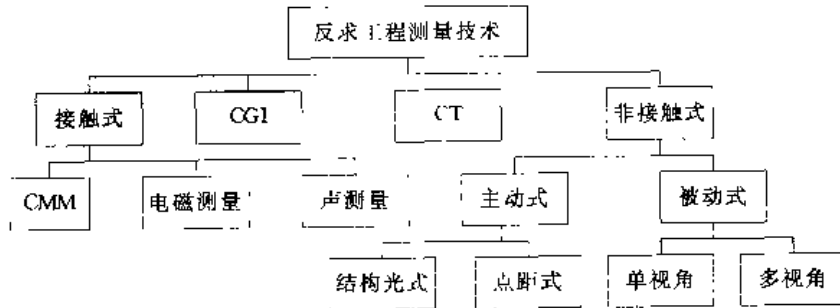


图 48.1-21 反求工程测量技术分类

接触式测量中,最常用的是采用机械式测头的三坐标测量机 CMM(Coordinate Measuring Machine)和有接触式滑动测头的三坐标测量仪。测头一般是灵敏的压力感应装置,通过记录在三个坐标轴上线性位移的大小,计算出各点的 X、Y、Z 坐标值。

非接触式光学测量可分为两大类:主动式测量和被动式测量。主动式测量采用特殊的光源,如激光、有特殊波长的可见光,而被动式测量则直接使用环境光。主动式测量可分为结构光式和点距式(Spot Ranging)。点距式的光源和感应元件布置在同一个轴上,通过计算入射光和反射光的相位差或时间差来确定测点的距离,并进一步确定其坐标。结构光式则使用三角原理来确定测点的坐标。结构光式根据光源的类型可进一步划分为:单射线、单光带、多光带,或者网格编码照射及摩尔拓扑的模式光。

反求测量系统最常用的主要以下几种,即:三坐标测量机(CMM)、工业 CT 机、核磁共振机(MRI)、激光扫描仪、自动断层扫描机(CGI)、摄影成像仪。

这几种测量设备用于反求工程时的测量精度、测量速度、测量对象、三维重构性和与快速成形设备的集成性比较如表 48.1-1 所示。

表 48.1-1 常用数据测量系统的某些性能指标比较

测量系统	所测数据精度	测量速度	测量对象	三维重构性	与快速成形机集成
三坐标测量仪	高	慢	测头可触及内外表面	差	差
工业 CT 机	低	较快	内外型腔	中	中
核磁共振机	低	较快	内外型腔	中	中
自动断层扫描机	中	中	内外型腔	好	好
激光扫描仪	中	快	外表面	好	好
摄影成像	中	快	外表面	好	中

三坐标测量机测量精度高,但测量速度较慢,一个复杂曲面的零件很难用三坐标测量机来逐层实现数字化,且受到测量头的限制,不易测量曲率变化大的凹陷处的数据;工业 CT 机和核磁共振机通过逐层扫描采集数据,由于这两种设备在断层法向的数据精度很差,层厚也在 1mm 左右,需要由专门的数据处理软件进行校正和插补,难以复制出精度较高的机械零件或模型;自动断层扫描机能测量物件的内外型腔数据,能完全重构出样件的几何形状,对复杂零件的整体复制特别有用,但必须以破坏原件为代价;激光扫描仪利用

激光束进行无接触快速地对物件的外表面进行测量,生成三维的点云数据,经数据拼合处理,再通过专门的数据处理软件能正确地重构物件的外部形状,这种测量方法由于采用激光束扫描测量,适合各种类型的工件,如软性物体表面,表面曲率变化大探头不易触及的物件表面及不允许磨损的物件表面。但无法测量其内部形状的数据,摄影成像是—种新的光学测量仪,在自然光环境下,利用光学三角形定理,由数码相机摄取物体表面的影像,经由数码影像处理器处理,生成点云数据,再由专门的点云数据处理软件进行对被测物体表面的三维重构。

激光扫描和摄影成像是尽管测量精度不及三坐标测量仪,但由于其高速的数据采集能力和相应的数据处理软件的配合,已愈来愈多地被应用于反求工程。

#### 4.1 光学测量设备介绍

目前光学测量设备从使用的光源分类,主要有两类,即激光和自然光。能提供此类测量设备的厂商主要有英国的 3D SCANNERS 公司、德国的 Kreon Industrie 公司和德国 GOM mbh 公司及 Steinbachler 公司。其产品的主要技术参数介绍如下。

##### (1) 3D SCANNERS REVERSA

表 48-1-2 3D SCANNER 的技术参数

测量头型号	REVERSA 25L	REVERSA 10H	REVERSA 25H	REVERSA 50H
应用对象	数学	珠宝、硬币测量	模型制作	产品设计、汽车设计
测量范围/mm	25	10	25	50
测量光管长度/mm	35	15	25	45
测量光管宽度/ $\mu\text{m}$	100	40	60	100
分辨率/ $\mu\text{m}$	25	5	10	20
重复精度/ $\mu\text{m}$	120	25	50	100
测量精度/ $\mu\text{m}$	50	10	20	40
最小测量间距/ $\mu\text{m}$	120	30	60	100
标准扫描速度/(点/s)	1000	1000	1000	1000
测量头重量/g	610	620	630	630
激光等级	3A	3A	3A	3A

##### (2) KREDN 激光测量系统

KREON 激光测量系统如图 48-1-24,测量头外形如图 48-1-25,其技术参数如表 48-1-3。

##### (3) ATOS 数据测量系统

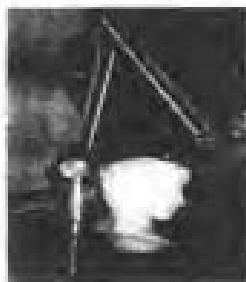


图 48-1-24 KREON 激光测量系统外形

英国 3D Scanners 公司的激光三维扫描仪 REVERSA 配有测量桌 REPLICAS90 的测量系统如图 48-1-22。

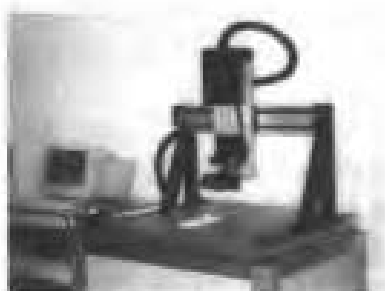


图 48-1-22 3D SCANNER 测量系统



图 48-1-23 3D SCANNER 测头  
3D SCANNER 测头形状如图 48-1-23,其技术参数如表 48-1-2。



图 48-1-25 KREON 测头

ATOS 测量系统的测量过程是基于光学三角形原理,系统将一组不同的光栅条纹投影到工件表面,通过两个 CCD 照相机分别从不同的角度对工件表面影像进行拍摄,然后将影像经数码影像处理,计算出每个像素点的三维

坐标值,最终获得工件表面的点云数据。

表 48-1-3 KREON 激光测量头技术参数

激光探头型号	KLS150	KLS170
单位获取区域上的点数	600	600
获取频率/(区域/s)	10	10
静止获取点速/(点/s)	达 15000	达 15000
量程/mm	50	120
平均激光线宽/mm	25	25
Z轴分辨率/ $\mu\text{m}$	$\pm 4$	$\pm 10$
Z轴精度/ $\mu\text{m}$	$\pm 10$	$\pm 20$
X和Y轴分辨率/ $\mu\text{m}$	45	120
重量/kg	1.905	2.220
应用环境温度/ $^{\circ}\text{C}$	达 40	达 40
电缆线长/m	5~30	5~30

ATOS 测量系统如图 48-1-26,测量头如图 48-1-27,技术参数如表 48-1-4。

表 48-1-4 ATOS 测量头技术参数

ATOS 测头型号	测头型号 1	测头型号 2	测头型号 3
测量尺寸/mm	380×230 ×280	190×145 ×140	12×7 ×8
测量精度/ $\mu\text{m}$	$\pm 7\sim 50$	$\pm 7\sim 25$	$\pm 7\sim 1.5$
焦距/ $\mu\text{m}$	500	250	16
记录时间/s	10	10	10
测量点数(标准 CCD 相机)	430296	430296	430296
测量头尺寸/mm	520×110 ×150	350×110 ×150	230×110 ×150
重量/g	2500	2300	2200
工作距离/mm	500	250	30

#### (4) COMET VariZoom 测量系统

COMET VariZoom 测量系统是一个可变焦系统,利用白光边缘投影法(使用投影网格和照相机的三角形测量法),物体表面可以通过非接触方式获得数字化点云数据。它是通过白光源将一系列格栅化光束投射到待测量的物体上,再用数码相机沿着光束方向将这些投射到物体表面上格栅进行拍摄。通过机械地移动格栅,投影的模式也会随之变化。因此,对每一帧在数码相机上获得的图片,都会分配有一个确定的编码。进



图 48-1-26 ATOS 测量系统

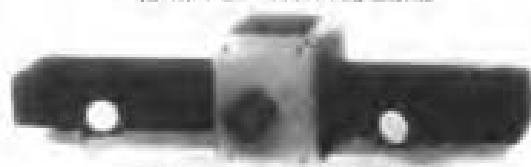


图 48-1-27 ATOS 测头外观

而,对物体每一个点的三维位置,可以从两个目标镜头—光源和照相机之间的距离 $b$ 及三角法中的角 $\alpha$ 和 $\beta$ 计算得到。由于在 COMET VariZoom 系统里内置有软件控制的调焦功能,用户可以方便地调整解像度,因此,对于物体表面上重要的局部细节就可以利用这种高解像度模式进行测量;而对于物体表面上大部分的同类曲面部分则可以采用标准模式进行测量;用户可以很容易地根据被测物体的特点制定最佳的测量方法,因而对于复杂的表面几何形状(凹状,凸状,桥形的等)用一部照相机就可完成。COMET 系统由以下四部分组成:①带控制器的光学测量头②支架系统③PC 机(Windows 操作系统)④COMET 数字化软件附加包。COMET VariZoom 系统配备有多种不同类型的测量头,软件模块和支架等。这就在保证数据质量的基础上,为不同的测量任务提供了很高的灵活性。该系统的主要技术指标如表 48-1-5。

表 48-1-5 COMET 测量头技术指标

标准 COMET 测量头的技术规格				
	C20	C75	C250	C400
测量范围/mm	45×35	78×58	280×180	430×330
分辨率/mm	$\pm 0.02$	$\pm 0.025$	$\pm 0.05$	$\pm 0.1$
测量距离/mm	350	400	820	820
测量头尺寸/mm	标准通用单元	275×164×124	440×164×140	440×164×140
COMET VariZoom 测量头的技术规格				
	C100 VZ	C250 VZ	C400 VZ	C250/400 VZ <sup>①</sup>
测量范围/mm	100×80	245×200	420×340	245×200
在高解像方案和标准 模式下				或 420×340
测量范围/mm	45×35	170×100	210×170	120×100
在变焦状态下				或 210×170
分辨率/mm	$\pm 0.02$	$\pm 0.04$	$\pm 0.07$	$\pm 0.04\sim 0.07$
测量距离/mm	430	820	820	820
测量头尺寸/mm	275×164×124	440×164×140	440×164×140	440×164×140

① 测量头模式 C250/400VZ 是一种简单的照相机的物镜转换特征表示,即在 C250VZ 和 C400VZ 测量头模式中转换。

### (5) T-Scan 手持式激光测量系统

该系统由一个多边形镜头定位的一根直线可视激光束(670nm, 激光等级 2), 通过高频(10kHz)扫描镜对物体表面进行扫描测量, 应用三角定律, 激光束在物体表面经反射后由激光接受器接收, 然后经计算获得物体表面的位置坐标。

图 48-1-28 光学跟踪器示意图。而激光扫描器的六度空间位置(三个空间定位加三个转角)则由光学跟踪器来确定。该跟踪器会自动捕捉激光扫描器上共 29 个红外定位点中至少三个点, 从而确定该激光扫描器的位置。

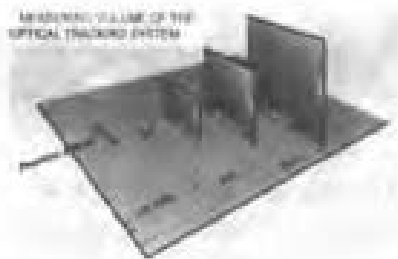


图 48-1-28 光学跟踪器示意

通过移动扫描器, 整个物体表面就能被记录下来。而测量下来的三维坐标点则会实时显示在计算机显示屏上。一个导航条用于确定在移动扫描器时能保持最佳的光束距离。

OMET T-Scan 扫描系统主要包括以下三个组成部分: 手持式激光扫描器、移动式光学跟踪器和基于移动架的计算机系统, 包括: COMET T-Scan 系统软件及针对激光扫描器和光学跟踪器的控制器, 如图 48-1-29 所示。

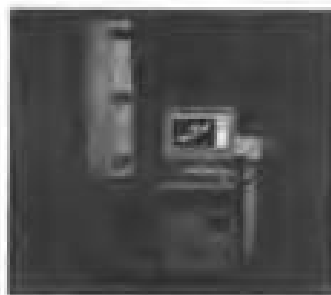


图 48-1-29 系统组成

其技术指标如表 48-1-6。

表 48-1-6 T-Scan 激光测量系统技术指标

测量深度	75mm	扫描头重量	1820g
扫描宽度	90mm	扫描头尺寸	185mm × 300mm × 145mm
平均测量距离	60mm	扫描头与计算机间电缆长度	9m
扫描频率	25~250kHz	分辨率	0.150mm
距离测量分辨率	1μm	激光类型	二极管

(续)

测量深度	75mm	扫描头重量	1820g
距离测量精度	+/- 20μm	波长	670nm
扫描方向点密度	0.30~0.5mm	激光等级	二级

## 4.2 重构软件介绍

### (1) 重构软件的基本功能介绍

这里介绍的重构软件是指专用于对点云数据进行曲面构建的软件。这类软件应具有的主要功能有:

1) 扫描点处理 对点云数据的预处理, 应包括如下功能:

对点云数据进行区域分割; 自动地把数以万计的扫描点细分为容易处理的区域。

截面: 可直接在点群上作不同类型割切(Sectioning), 所得切点可自动变为曲线。

过滤: 各种数据减少和修匀(Smoothing)过滤器, 包括抽样、平均、高斯型(Gaussing)、弦度偏差(Chordal deviation)和角度偏差(Angular deviation)。

特征点抽取, 特征点抽取器(Feature Extraction), 可在庞大的数据中识别出各种特征或主轴线。

2) 曲线处理 构造曲线和对曲线进行编辑的工具, 应具有以下功能:

曲线生成: 在规定的公差范围内把扫描点直接变成 NURBS 曲线。

曲线修改: 减少多余控制点(Control points), 修匀(Smoothing), 延长(Extend)和作即时交换(Interactively)形状修改。

曲线投影: 投影曲线至扫描点群或曲面上。

曲线连接: 将不同曲线缝合, 且保持或产生连续性(Continuity)。

3) 曲面构建工具 提供各种构建曲面的方法, 包括:

由扫描点生成曲面, 可由扫描点直接产生曲面而不需经过先构造曲线的过程, 也可先构造周边曲线再用其内点群产生曲面。

由曲线产生曲面, 功能包括混合(Blending)3 或 4 条 NURBS 曲线, 拉伸(Lofting)多曲线, 旋转(Revolving)曲线, 或扫描(Sweeping)曲线沿一条或两条导线(Guide curves)。

曲面生成而其特征保留, 拉伸多曲线生成曲面而能保留曲线上的特征。

由二次方程生成曲面, 由二次曲线生成圆锥面、圆柱面和球面。

4) 曲面处理工具 其功能包括:

连接曲面: 在两张分开的曲面之间产生第三张连

接且连续的曲面。

**混合曲面:**将 2、3 或 4 张曲面混合 (Blend and Stitch), 且保持或产生连续性。

**曲面修改:**处理表面上的控制点 (Control points) 做交互式形状修改。

**裁剪曲面:**裁剪 (Trim) 曲面于指定的曲线内。

**变半径倒圆角:**在两曲面间产生相等或变半径倒圆角 (Constant or Variable radius fillet)。

**将曲面贴近点群:**自动改变现有曲面来配合点群的形状, 且保持其连续性。

**曲面修匀:**光顺 (Smoothing) 曲面且保持与连接曲面的连续性。

5) 曲面分析工具 其功能包括:

**点群与曲面之间的距离比较:**可找出在点群中的每一点与所构造曲面之间的实际距离。

**横截面分析:**在选择曲面上显示任意截面 (Sections)。

**反射:**在被照射的模型上显示等反射率线 (Iso-reflectance contours)。

**曲率:**根据高斯型曲率计算 (Gaussian curvature) 来显示等曲率线 (Iso-curvature contours)。

**连续:**显示曲面与曲面之间的连续性 (Degree of continuity)。

6) 系统的兼容性。三维重构软件应具有连接各种 CAD/CAM, 有限元分析软件和快速成形 (Rapid Prototyping) 系统的能力, 有多种输入/输出数据格式。

## (2) ICEM Surf 商用软件介绍

ICEM Surf 是 ICEM 公司的产品, 用于构建不规则形状曲面的工具, 达到 A 类曲面处理的能力。其动态曲面诊断工具可对曲面制造可行性进行评估。其独到的曲面造型能力将美学与产品设计造型和工程融为一体, 是当今这一领域的优秀曲面造型软件。同时, ICEM Surf 与 Pro/ENGINEER 通过 ATB (相关拓扑总线), ICEM Surf 和 Pro/ENGINEER 的格式可以互相使用。并且还提供与 CADD5、I-DEAS Master Series、CATIA、Unigraphics 和 AutoForm 等主要 CAD/CAM 系统的接口。

ICEM Surf 的软件模块可根据需要选配, 它包括以下主要模块:

ICEM Surf Professional;

ICEM Surf Magic;

ICEM Surf Master;

ICEM Surf Scan;

ICEM Surf Clay。

ICEM Surf Professional 是其基本模块, 用于建立曲线、曲面、分析和修改设计并接收动态曲面诊断的实

时反馈。

ICEM Surf Magic 是使用实时整体建模技术和一整套渲染和可视化工具, 用来设计复杂的数据对象并为其建模。它可以自动调整和修改现有的曲面模型, 从整体上有效地适配重新数字化过的模型, 可以快速地展现对模型所做的更改。

ICEM Surf Master 是整体偏移工具, 能够用原始曲面建立具有恒定间隙的偏移曲面, 其中包括保存转换后曲面质量和修整或外推偏移曲面。

ICEM Surf Scan 用于处理点云、分成三角形、过滤、偏移、诊断并修改数据。用户可以按照要求, 组织和整理数据, 建立剖面并提取特征线, 从而允许建立 ICEM Surf 特殊曲面。它的输入格式有 ESF、STL、ICEM CFD/CAE 域文件、AutoForm、ATOS-GOM、Steinbichler AC 和 Hymarc。输出格式有 STL、ICEM CFD/CAE 域文件和 AutoForm。

ICEM Surf Grid 该模块提供了一个自动的曲面三角网格生成器或在曲面模型或 STL 模型上建立四面网格的可选功能。产生的网格可以被用于计算流体动力学分析, 或用于快速原型设计应用。

ICEM Surf Synergy 这个模块将允许用户与远程用户进行在线会话。两个用户 (本地和远程) 都必须有这个模块的使用许可证, 他们两个都必须使用 NT 平台或 UNIX 工作站, 而且, 他们必须具有在本地得到许可的相同的 ICEM Surf 版本。

ICEM Surf Clay 该模块允许小面模型的局部和整体建模 (“虚拟陶土建模”)。它的功能可以被应用到概念设计、可用性检查、工模具设计等。

ICEM Surf Rendering ICEM Surf 渲染以 LightWork 设计技术为基础, 提供了最优渲染。该模块能够在 ICEM Surf 中建立高级渲染和逼真图像。

ICEM Surf 转换器, 它提供多种的数据交换接口, 如:

ICEM Surf 直接接口 ICEM Surf 直接接口是直接数据交换工具, 用于从曲面处理到 Cax 系统: CATIA、CADD5、I-DEAS Master Series、Unigraphics 和 AutoForm。ICEM Surf 专业版包括了 Pro/ENGINEER (通过 ATB) 和 CDRS 的直接转换器。

ICEM IGES ICEM IGES 转换器在 ICEM CAD/CAM 系统和多个其他 CAD/CAM 系统之间转换图形数据。ICEM VDA/FS ICEM VDA/FS (Verband der Automobilindustrie) 曲面接口是不同销售商的应用系统之间交换曲面几何体数据的工业标准。

SET SET (Standard d'Echange et de Transfert) 输入和输出由 ICEM Surf 支持。这是法国汽车和航空市场交换几何体数据常用的一个标准格式。

## 5 应用实例

### (1) 砂轮型面扫描

通过对砂轮型面的测量,分析砂轮切削刃的磨损与被加工零件表面粗糙度的关系,见图 48-1-30。

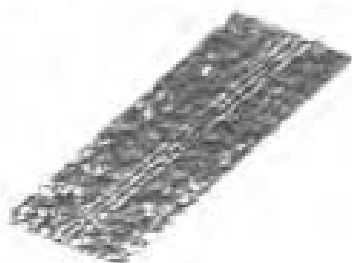


图 48-1-30 砂轮型面扫描点云数据

### (2) 自行车反光块各斜面倾角测量

实物:自行车反光块

用途:测量各斜面反光角度

技术方案:根据测量数据重构出斜面及基准面,计算其夹角

技术特点:精密测量,统计学方法的应用  
反光块扫描点云数据见图 48-1-31。

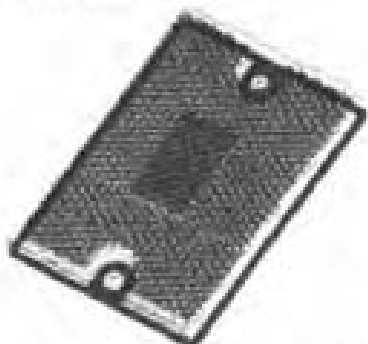


图 48-1-31 反光块扫描点云数据

### (3) 汽车车灯反光面扫描并重构

实物:三种型号的汽车车灯(见图 48-1-32~48-1-34)

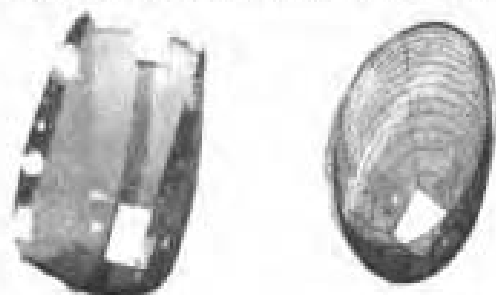


图 48-1-32 车灯反光面一 图 48-1-33 车灯反光面二

用途:借鉴车灯反光面型面形状,进行产品设计

技术方案:根据测量数据重构反光面

技术特点:对曲面性能要求高,可进行几何分析

(4) 某汽车研究所和汽车公司的后车灯设计



图 48-1-34 车灯反光面三

### 后车灯及车身侧围的配合

实物:后车灯及车身侧围(见图 48-1-35~图 48-1-37)

目标:解决干涉问题,改进型面设计

技术方案:第一步,扫描并重构两个型面,诊断干涉部位及原因;

第二步,修改曲面,解决问题

技术难点:基准的测量与其他部分的 CAD 数据配合

① 用 3D SCANNER 扫描;② 根据点云数据拟合出曲线;③ 由点云和曲线拟合出曲面。

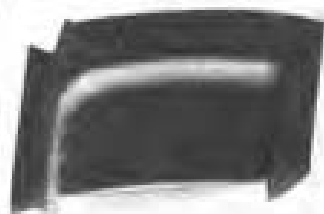


图 48-1-35 车身侧围扫描点云数据

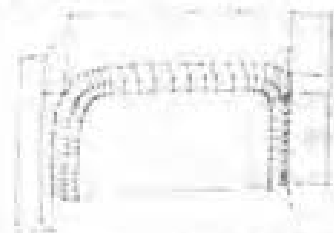


图 48-1-36 车身侧围曲线模型

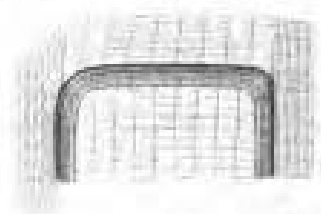


图 48-1-37 车身侧围曲面模型

实物:后车灯(见图 48-1-38~图 48-1-39)

目标:改进曲面形状,修复模具

技术方案:第一步,用油泥修改原零件

第二步,扫描并重构出曲面

第三步,修改模具



图 48-1-38 后车灯外表面  
扫描点云模型

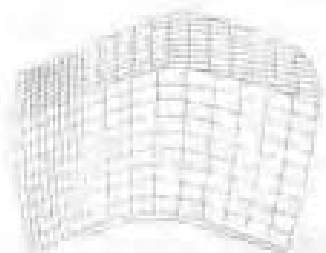


图 48-1-39 后车灯外表面  
曲面模型

(5) 扫描、重构并再设计汽车后视镜  
实物:汽车后视镜(见图 48-1-40—图 48-1-41)  
目标:借鉴原产品,实现新产品设计  
技术方案:第一步,扫描出完整的实物表面数据

第二步,进行曲线、曲面拟合

第三步,改进曲面,进行创新设计

技术流程:

1) ATOS 测点云数据

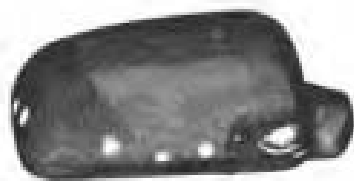


图 48-1-40 汽车后视镜  
扫描点云数据

2) 区域划分及曲线拟合

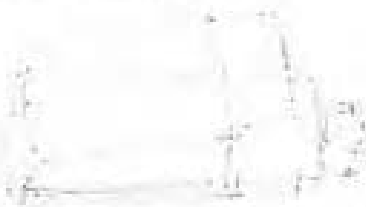


图 48-1-41 汽车后视镜的区域  
划分及拟合出的曲线模型

3) 重构曲面

4) 实体 STL 化

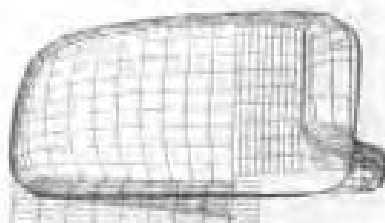


图 48-1-42 汽车后视镜的曲面模型

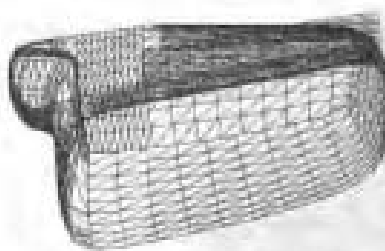


图 48-1-43 汽车后视镜三角化后的 STL 模型

## 6 反求设计全过程实例

下面以某汽车仪表盘为例说明反求设计的全过程:

1) 图 48-1-44 为汽车仪表板的样件,在此样件上进行定位球的粘贴(也可用其他定位方法)以便在点云处理和拼接时作为拼合的参考,为防止点云数据采集时的光线干扰,需要对容易反光的表面作防反光处理的操作。

2) 经上述处理后的仪表板样件进行激光或其他方式的扫描,根据样件的形状,移动样件或改变扫描头的位置对需要的表面行数据采集,通常需获取多幅数据点云,经去除噪声和根据定位球将多幅点云数据作拼合使得仪表板完整的表面外形点云数据,如图 48-1-45 所示。



图 48-1-44 汽车仪表板样件

3) 对点云数据进行边界和特征线提取,曲面片分割,然后进行曲面重构,使重构后的曲面与点云的误差在规定的范围内。经曲面重构后的表面模型如图 48-1-46。

4) 在曲面模型的基础上进行实体建模,按仪表盘上安装的仪表尺寸和固定位置对仪表盘进行结构设计,完成对仪表盘零件的再设计,图 48-1-47 是用 CATIA 软件完成的实体模型。





图 48-1-45 汽车仪表板的点云数据

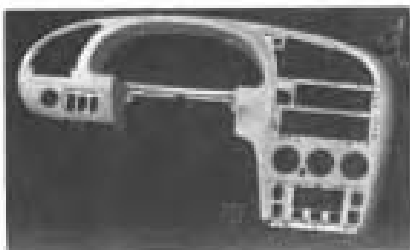


图 48-1-46 汽车仪表板的曲面模型

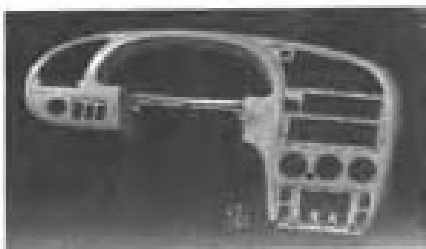


图 48-1-47 汽车仪表板的实体模型

5) 对完成的实体模型与设计要求进行误差分析, 针对误差对仪表板进行再设计, 局部调整实体的形状和尺寸, 使最终的形状尺寸满足设计要求。图 48-1-48 为误差分布情况。

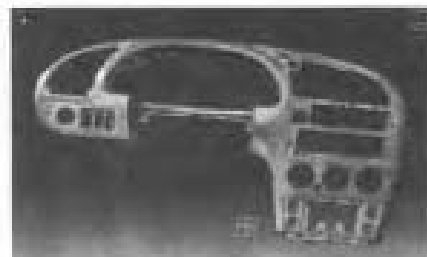


图 48-1-48 仪表板的误差分布情况

6) 最终完成的模型, 根据实际需要可转换成其他格式的模型, 便于加工。图 48-1-49 为模型转换界面。



图 48-1-49 实体模型转换界面

## 第2章 快速成形

### 1 快速成形设备发展简况

受三维形体二维处理思维的启迪下,与传统的去除材料成形相反,用材料累加成形的——一种全新的形状加工技术——快速成形(Rapid Prototyping,简称RP)应运而生,这一技术的出现被认为是近代制造技术领域的一次重大突破,其对制造业的影响可与当年的数控技术出现相媲美,RP系统综合了机械工程、CAD、数控技术、激光技术、自动控制技术和材料科学技术,是一项综合交叉技术的结果。它可以自动、直接、快速、精确地将设计思想物化为具有一定功能的原型甚至直接制造零件,从而可以对产品设计进行快速评估、修改和部分的性能试验,可有效地缩短产品研发周期,快速提供市场需要的产品。这主要体现在模具设计和开发的一次成功率提高,避免由于设计的疏忽造成模具返工的损失,利用快速成形件装配的产品进行客户接受性评估,以避免由于产品的滞销带来的资金浪费。在其他领域快速成形技术的应用也愈来愈多,从而引发了快速成形设备的开发。从20世纪80年代末首先在美国开发成功第一台商用的快速成形机后,快速成形机产品不断面市,具有代表性的产品推出时间如下:

1987年,美国3D System公司,推出Stereolithography(SL)工艺的SLA-250原型机。

1988年,日本NTT Data/CMET公司,推出Solid Object Ultraviolet Plotter即SOUP机型。

1991年,美国Stratasys公司,推出Fused Deposition Modeling(FDM1650);美国Helisys公司,推出采用激光切割纸材的Laminated Object Manufacturing(LOM)系统。

1992年,美国DTM公司,推出采用粉末烧结技术的Selection Laser Sintering(SLS)系统。

1994年,快速成形机的开发达到了高潮,如美国的Sanders公司推出采用喷黑打印头技术的Model Maker系统;日本KiraCorp公司推出采用普通纸的Non-Stereolithography System;德国Fockele & Schwarze公司推出SLA机;德国EOS公司推出采用SLS工艺的EOSINT。

1996年,美国BPM公司推出用蜡质材料的喷黑打印喷头喷射成形Personal Modeler 2100机型;新加坡Kinergy Zippy公司推出LOM机,美国Stratasys

推出用ABS块状材料的Genisys。

1999年美国3D System推出使用352个喷头Theramojet Solid Object Printer和SLA-700。1994年后,华中科技大学、西安交通大学、清华大学和北京隆源公司也分别推出LOM、SLA和SLS的快速成形机,对我国的快速成形技术的研究和推广起了积极的作用。

### 2 快速原型技术流程

快速成形是近十几年来形成的一种全新加工技术,它完全摆脱了传统的“去除材料”加工方法,而是采用全新的“增加材料”加工法,将材料一层一层叠加的方法完成零件的加工,其方法是将复杂的三维加工分解成简单的二维加工的组合。因此,它不必需要传统的加工机床、刀具和工模具,也不受零件形状的复杂程度的影响,就能直接按计算机设计的三维实体模型制造出产品样品。

快速成形技术的基本原理是:将计算机上完成的产品三维设计模型,对其进行分层切片,得到各层截面的轮廓,按照这些轮廓,激光束选择性地切割一层一层的纸或固化一层层光敏树脂或烧结一层层的粉末材料,或者用喷射头选择性地喷射一层层热融材料或粘结剂,形成各截面轮廓并依此一层一层地叠加成三维产品形状。其过程如图48.2-1。

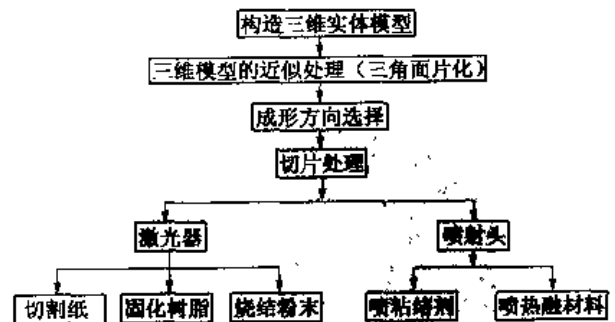


图48.2-1 快速成形流程示意图

#### (1) 三维实体模型的构建

按材料添加原理制造的快速成形系统,要求在CAD软件生成的三维实体模型基础上,对其进行分层切片,生成二维图形,按成片输入系统处理。通常,首先应在高性能个人计算机或工作站上,利用计算机辅助设计软件(如UGII、PRO/E、I-DEAS、SOLIDWORKS等)或用扫描机对已有的产品进行扫

描重构三维模型,根据产品的应用要求设计三维模型,或者根据产品的二维图转换成三维模型,图 48-2-2 是经激光三维扫描获得的点云形式三维模型。

### (2) 三维实体模型的近似处理

由于产品上往往有一些不规则的自由曲面,在加工前必须对其进行近似处理。目前的快速成形系统中,最常见的近似处理方法是采用一系列小三角形平面来逼近自由曲面,其中,每个三角形用 3 个顶点坐标和一个法向量来描述,如图 48-2-3。三角形的大小可按用户的要求进行选择,从而能得到不同的曲面近似精度,经这样近似处理的三维模型文件称为 STL 格式文件,它由一系列彼此相连的空间三角形组成。图 48-2-4 是某照相机镜架的 STL 表示。目前的商品化三维建模软件一般都有将三维模型转换成 STL 格式文件的功能,但由于建模的方法因人而异,有时会隐含一些不易发现的错误,输出的 STL 文件会有一些错误,造成成形时出现成形路径出错,需要经软件自动检测或人工逐层检查,局部修补。



图 48-2-2 某照相机镜架的扫描三维模型

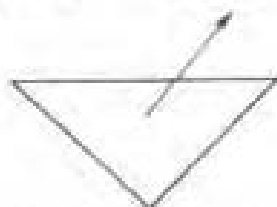


图 48-2-3 三角形的表示



图 48-2-4 照相机镜架的 STL 表示

### (3) 成形方向选择

成形零件的精度、表面质量及成形所需时间与零件的成形方向有密切关系,也涉及零件的制造成本,用最低的成本制造符合要求的零件,必须综合考虑零件的成形方向。这是因为不同的成形方向生成的支撑不同,被支撑的面不同,因而,支撑的生成时间和所消耗

的支撑材料也不一样。

### (4) 切片处理

因为快速成形是按切层的截面形状来进行加工,所以,加工前必须将三维模型按选定的成形高度方向,每隔一定的间隔进行切片处理以便提取截面的形状,间隔的大小根据被成形件的精度和生产率的要求选定,间隔愈小,精度愈高,成形所花的时间愈长,反则亦然。通常,间隔的范围为 0.05~0.5mm,常用 0.1mm 左右,在此取值下,已经能得到相当光滑的成形表面。切片间隔选定后,成形时每层叠加的材料厚度应与其相适应,因为成形高度方向的尺寸精度完全由层厚所决定。商品化的各种快速成形系统均带有切片处理软件,具有自动提取每层截面的形状和编辑的功能。

### (5) 截面形状和三维实体的加工

经切片处理后的截面形状加工方式,根据不同成形原理制造的设备,成形方法也不相同。主要有两种,即按轮廓和面积。按轮廓方式加工,是在经预处理的纸或其他材料薄片上,采用激光沿截面的边界切割,按切片处理的高度方向,切割一层粘合一层,一层一层进行叠加而加工成三维实体。按面积方式加工,是根据截面的形状和厚度,采用类似数控平面加工方式将被加工材料固化或粘结(通常采用激光)或喷射液滴状态的成形材料堆积成截面,按切片处理的高度方向一层一层地堆积成三维实体。

### (6) 表面处理

表面处理是将成形后的实体进行后处理,后处理主要是对生成的原型在保证精度和形状的前提下,对其表面进行光整处理,主要的方法有磨光、填补、粘结(当受成形空间限制,需分割成形的情况)、喷涂等。对原型件的后处理是快速成形的重要工序,根据工件的不同形状和复杂程度,后处理的方式和打磨工具都有很大的差别,其中有很多靠经验积累。

## 3 几种典型的快速成形技术介绍

### (1) 光固化立体造型 (SLA, Stereolithography Apparatus)

该技术以液态光敏树脂为原料,将计算机控制下的激光束按预定零件各分层截面的轮廓为轨迹对液态光敏树脂逐点扫描,使被扫描区的树脂薄层产生光聚合反应而固化,从而形成零件的一个薄层截面。当一层固化完毕,移动工作台,在原光固化好的树脂表面上再敷上一层新的液态光敏树脂,以便进行下一层的扫描固化,新固化的一层牢固地粘在前一层上,如此重复进行直至整个零件原型制造完毕。简单的原理示意图如图 48-2-5 所示。

### (2) 选择性激光烧结 (SLS, Selected Laser

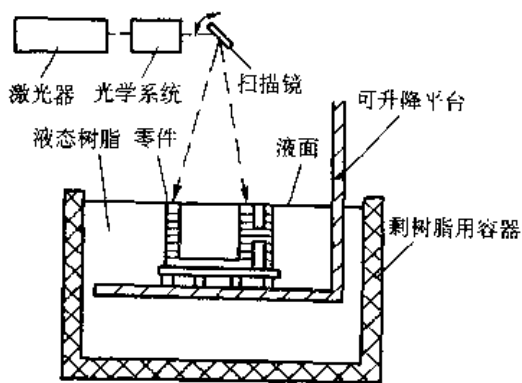


图 48.2-5 SLA 工作原理示意图

Sintering)

选择性激光烧结是利用红外激光束的热能熔化热塑性材料以形成三维零件。目前该方法采用 CO<sub>2</sub> 激光器作能源,各种粉末材料为造型材料。在工作台上均匀铺上一层很薄的粉末(100~200μm),激光束在计算机控制下按照零件分层轮廓有选择性地地进行烧结,一层烧结完了,铺粉液筒在机构的控制下重新补上一层薄的粉末,再进行下一层烧,直到整个零件烧结完毕,然后去掉多余的粉末,再进行打磨、烘干等处理,便获得零件。这种成形的工作原理示意图如图 48.2-6 所示。

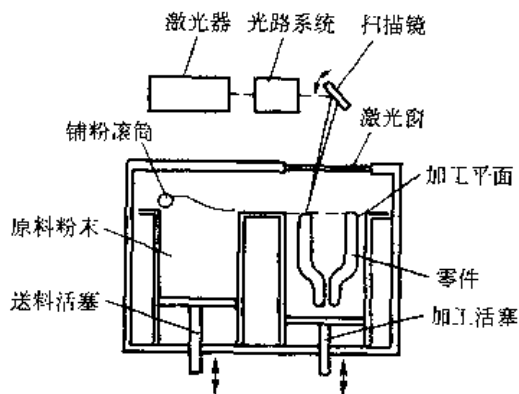


图 48.2-6 SLS 工作原理示意图

(3) 熔融沉积造型 (FDM, Fused Deposition Modeling)

FDM 是利用热塑性细丝材料在移动的喷头中进行熔化,熔化后的材料在喷头的运动过程被挤压出来堆积零件。FDM 工艺的关键是保持半流动成形材料刚好在熔点之上(通常控制在比熔点高 1°C 左右)。FDM 喷头受 CAD 分层数据控制使半流动状态的熔丝材料(丝材直径一般在 1.5mm 左右),从喷头中挤压出来,凝固形成轮廓形状的薄层。每层厚度范围在 0.025~0.762mm,一层叠一层最后形成整个零件模型。图 48.2-7 为 FDM 工作原理示意图。

(4) 叠层制造 (LOM, LAMINATED Object Manufacturing)

逐层物体制造技术是通过逐层激光切割薄纸材料制作零件的一种技术。LOM 工艺将单面涂有热熔胶的纸片通过加热辊加热粘接在一起,位于上方的激光器按照 CAD 分层模型所获数据,用激光束将纸切割成所制零件的内外轮廓。然后新的一层纸再叠加在上面,通过热压装置和下面已切割层粘合在一起,激光束再次切割,这样反复逐层切割—粘合一切割直至整个零件模型制作完成。图 48.2-8 为 LOM 工作原理示意图。

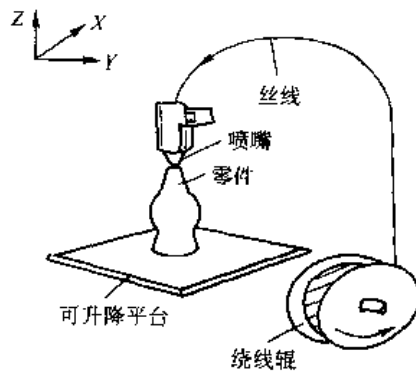


图 48.2-7 FDM 工作原理示意图

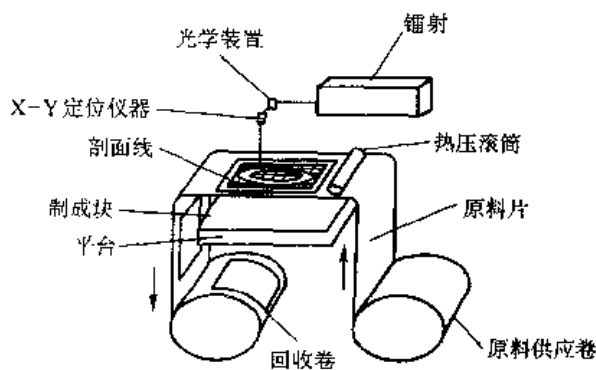


图 48.2-8 LOM 工作原理示意图

## 4 快速成形技术的应用

(1) 快速成形技术应用发展概述

快速成形 (Rapid Prototyping) 技术是 20 世纪 80 年代末,由美国科学家研制成功,1988 年,美国 3D SYSTEM 公司推出了世界上第一台商品化快速原型机 SLA-250。由于人们看到它的前景和活力,许多发达国家如欧洲、日本,从无到有也相继开发出了自己的商用快速原型机,1992 年以前全世界总共装机为 300 台,而到 1997 年世界装机量为 1400 台。分布于六大洲的 40 多个国家,这期间几乎以每年 50% 的速度增长,而从 1995 年到 1996 年 4 月,据美国 DAYTON 国际 RP 技术研讨会上的一份报告统计台数增长为 40% 以上,而在新产品设计开发、制造中使用过 RP 技术的厂

商数目前增长率为 52%。同时,为了满足中小企业对该技术的要求,从 1992 年起开始在世界各地建立起了 RP 技术服务中心。1992 年建立的服务中心有 42 家,而到 1999 年已增加到 350 家,其中主要分布在欧洲和美国。国内在 RP 领域的研究始于 1995 年,1995 年国家科委将 RP 技术正式列为国家“九五”攻关项目,同年 12 月在清华大学召开了国内首次 RP&M 会议,与会人数只有 30 人左右,主要是来自高校和科研单位的研究人员,而开展 RP 应用研究和服务的只有广州殷发公司一家,在此后的两年多时间里,由于人们看到了 RP 技术的应用前景和给工业界带来的巨大效益,许多高校和科研单位加入了 RP 技术开发和研究的行列。同时,也开始研究国产商用的快速原型机。西安交通大学、清华大学、北京隆源公司、华中理工大学在这方面都开展了研究,并有了自己的商用产品。同时国内的一些企业和公司也开始从国外引进 RP 设备,其中有广东华泰公司引进 SLA500 和 FDM 机,西安工业技术交流引进 SOUP500,西安交大引进 LOM 机,清

华大学引进 SLA-250,新乡新飞冰箱厂引进 SLA-500,上海汽车研究所引进 SOUP600,华宝、康佳、科龙、春兰、海尔、海信、上海德尔福等也都购买了快速成形设备,用于加速引进产品的国产化进程和新产品开发。但从量大面广的中小企业,从目前的能力和技术力量都存在一定的困难,需要有社会力量的帮助才能把这一技术引入产品开发,依靠政府推动和高校、研究所的技术支持促进和帮助中小企业应用快速成形新技术提升产品的能级和市场的竞争力。

快速成形技术由于其独特的自由成形概念,引起了机械制造的一场革命,它尽管问世不长,但给制造业带来了巨大的效益,从而促使这一技术应用的日益广泛。据 1996 年的统计,全世界共有快速成形系统 2243 台(大部分分布在美国和欧洲),仅 1996 年的系统销售量达 763 台,用快速成形系统服务的机构有 284 个,总产值近 5 亿美元。

## (2) 快速成形技术的主要应用

快速成形技术的主要应用如图 48.2-9。

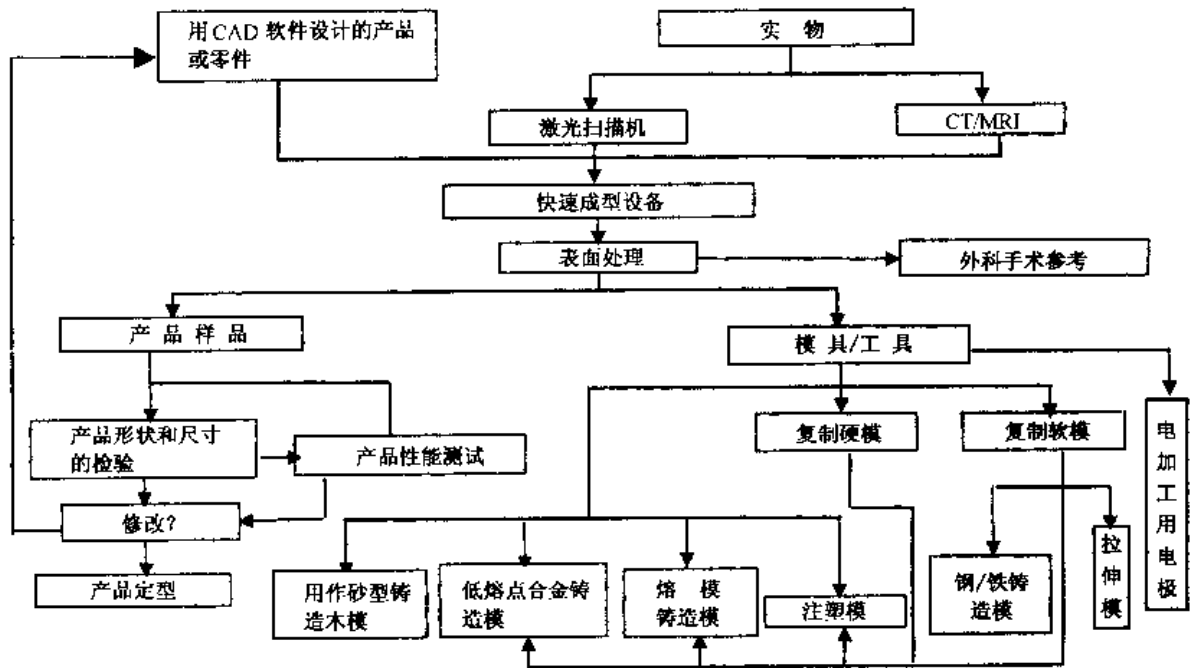


图 48.2-9 快速成形技术的主要应用

1) 制造产品样件 用快速成形系统直接制造产品样件,一般只需传统加工方法 30%~50% 的工时。这种样件与最终产品相比,虽然在材质方面有所差别,但在形状及尺寸方面几乎完全相同,而且有较好的机械强度,经适当表面处理(如表面喷涂金属或油漆)后,其外观与真实产品完全一样,因此,可用于给设计者和用户对产品进行直观检测、评价和制作产品样本,最大限度地获取市场对产品的反馈意见,并可迅速地反复

修改,以获得最大的使用和市场价值。

2) 产品性能的测试、校验和分析 用快速成形系统直接制造的产品样件,可对单个零件和装配件的加工工艺性能、可装配性和相关的工模具的校验与分析,还可用于运动特性的测试、风洞试验、有限元分析结果的实体表达等。

据 KINERGY 公司的有关资料介绍,美国的 Sundstrand Aerospace 公司设计的飞机用发电机,有

很大的箱体和装于其内的1200多个零件构成,仅仅箱体的设计图纸就有50多张,至少有3000多个尺寸,因为箱体的形状太复杂,不仅使设计校验十分困难,而且给制造铸模造成很大的麻烦,需耗时3~4个月,费用高达8万美元。后来,公司采用快速成形技术,仅花2周就获得了产品样品,6周就做出了铸造用砂型。据此样品进行对该发电机的形状、尺寸、装配关系和部分功能校验以及机械加工工艺设计、工装模具校验和装配顺序设计,获得了很大的成功。

### 3) 工装模具的制作

① 直接制作模具 由于快速成形样件有较好的机械强度和稳定性,经表面处理(如喷涂清漆、高分子材料或金属)后,可直接用作某些模具,如砂型铸造的木模、低熔点合金铸造模、试制用注塑模以及熔模铸造的蜡模的替代模或蜡模成形模。

② 用快速成形样件作母模,复制软模具 用快速成形样件作母模,可浇注蜡、硅橡胶、环氧树脂、聚氨酯等软材料,做成软模具;或先浇注硅橡胶、环氧树脂模(即蜡模的成形模),再浇注蜡模。蜡模用于熔模铸造,环氧树脂模用作试制用注塑模或低熔点合金铸造模。

③ 用快速成形样件作母模,复制硬模具 用快速成形样件作母模,或根据其复制的软模具,可浇注(或涂覆)石膏、陶瓷、金属基合成材料、金属,做成硬模具。从而批量生产塑料件或金属件。如用金属基合成材料浇注成的蜡模的成形模,模具寿命可达1000~10000件。

④ 制作电脉冲机床用电极 用快速成形样件作母体,通过喷镀或涂覆金属、粉末冶金、精密铸造、浇注石墨粉或特殊研磨,可制作金属电极或石墨电极。

4) 外科手术的辅助手段 快速成形系统可利用CT扫描或MRI核磁共振的图像数据,制作人体器官模型,如头颅、面部或牙床,供外科提供对复杂手术的操作,为骨移植设计样板或将其作为X光检查的参考手段,提高手术的成功率。

5) 用快速成形技术制作注塑模 用快速成形技术制作注塑模,是制造业最关心的一种应用。通常,注塑模由型腔、浇注系统和冷却系统等组成,其中,型腔部分最复杂,加工最困难,可用快速成形法制造;而浇注系统和冷却系统部分为标准结构,可另嵌入,构成组合模。用快速成形法制造注塑模型腔的主要方法和工艺过程如图48.2-10所示。

① 直接用快速成形的纸基模构成形腔 用KINERGY公司ZIPPY系列快速成形系统制造的纸基模有较好的力学性能,经表面处理(如喷镀)后,可用作试制用注塑模型腔。常用的喷涂方法有金属电弧

喷镀、金属基合成材料涂覆或陶瓷基合成材料涂覆。其中,金属电弧喷镀的材料为铝锌合金,喷镀厚度可达3mm。金属基合成材料又称为“液态金属”,在室温下呈胶体状,16h后固化,常用的有铝基合成材料,它是铝粉、环氧树脂和粘结剂的混合物,其抗压强度为70~80MPa,工作温度可达140°C,固化收缩率约为0.01%。陶瓷基合成材料的抗压强度为251.7MPa,工作温度可达260°C。

② 用快速成形样件作母模,复制软型腔 用快速成形样件作母模,可浇注硅橡胶、环氧树脂等软材料,构成试制用注塑模的软型腔。例如,TEK高温硫化硅橡胶的抗压强度可达12.4~62.1MPa,工作温度可达150~500°C,模具寿命可达200~500件。一般室温固化硅橡胶构成的型腔寿命为10~25件。

③ 用快速成形样件作母模,复制硬型腔 用快速成形样件作母模,据其复制的软模,可浇注石膏、陶瓷、金属基合成材料、金属,构成硬型腔,从而批量生产塑料件。这些材料有良好的机械加工性能,可进行局部切削加工,以便获得更高的精度,或嵌入浇注系统和冷却系统等。其方法如下:

#### a. 用熔模铸造法制造铁型腔

只需制造1件钢或铁型腔时:

用快速成形系统制作纸质母模→浸母模于陶瓷砂液,形成模壳→在炉中固化母壳,烧去母模→在炉中预热模壳→在模壳中浇铸钢或铁型腔→型腔表面抛光→加入浇注系统和冷却系统等→批量生产用注塑模。

需制造多件钢或铁型腔时:

用快速成形系统制作纸质母模→用金属表面喷镀,或铝基合成材料、硅橡胶、环氧树脂、聚氨酯浇注法,构成蜡模的成形模→在成形模中,用熔化蜡浇注蜡模→浸蜡模于陶瓷砂液,形成模壳→在炉中固化模壳,溶化蜡模→在炉中预热模壳→在模壳中浇铸钢或铁型腔→型腔表面抛光→加入浇注系统和冷却系统等→批量生产用注塑模。

其中蜡模的成形模可反复使用,以浇注多个蜡模,从而制造多件钢或铁型腔。

#### b. 用铝基合成材料浇注型腔

用快速成形系统制作纸质母模→浇注硅橡胶、环氧树脂、聚氨酯等材料,构成软模,移去母模,在软模中,在室温下用铝基合成材料浇注型腔→涂覆陶瓷基合成材料型腔表面抛光→加入浇注系统和冷却系统等。

小批量生产用注塑模。其中插入软模的步骤是为了容易脱模。

#### c. 用陶瓷或石膏膜浇注注塑钢或铁型腔

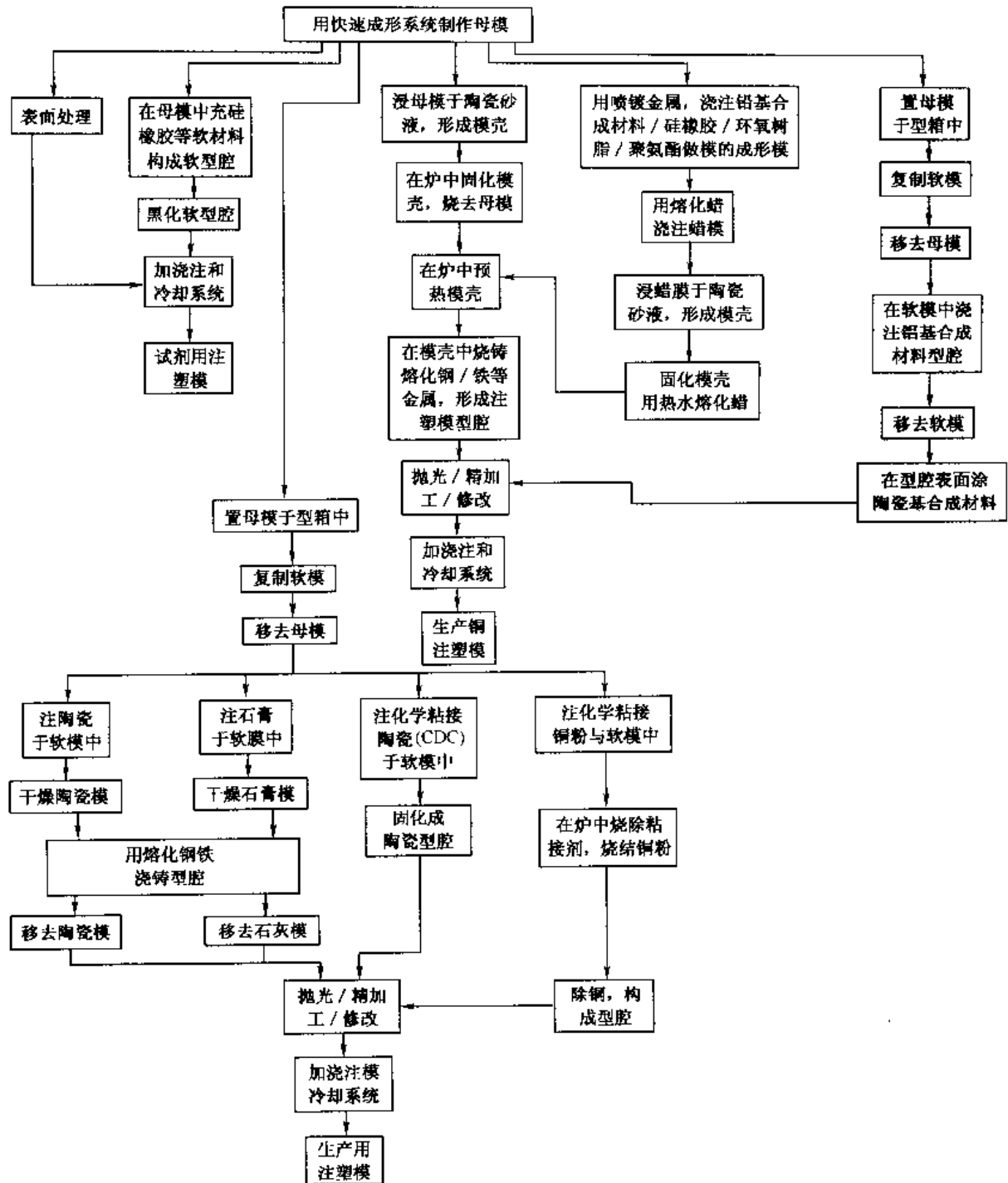


图 48.2-10 用快速成形法制造注塑的型腔的主要方法和工艺过程

用快速成形系统制作纸质母模→浇注硅橡胶、环氧树脂、聚氨酯等软材料→构成软模→移去母模→在软模中浇注陶瓷或石膏模→浇注铜或铁型腔→型腔表面抛光→加入浇注系统和冷却系统等→小批量生产用注塑模。

d. 用快速成形系统制作纸质母模→浇注硅橡胶、环氧树脂、聚氨酯等软材料→构成软模→移去母模→浇注化学粘结陶瓷(EDC, 陶瓷基合成材料)

型腔→在 205℃→下固体 CBC 型腔→型腔表面抛光→加入浇注系统和冷却系统等→小批量生产用注塑模。

这种化学粘结陶瓷型腔的寿命为 300 件。

e. 用快速成形系统制作纸质母模→浇注硅橡胶、环氧树脂、聚氨酯等软材料→构成软模→移去母模→在软模中浇注化学粘结铜粉型腔→在炉中烧除型腔内的粘结剂, 烧结铜粉→型腔渗铜→型

腔表面抛光 → 加入浇注系统和冷却系统等 → 批量生产用注塑模。

6) 用快速成形技术制作电脉冲机床用电极 用快速成形技术制作电脉冲机床用电极,也是制造业十分关切的一个应用方面,通常有如下方法:

① 喷镀、涂覆法 用快速成形系统制造母模 → (用母模烧铸石膏模 → 在石膏模中)用金属喷镀(或涂覆、电镀、电泳)2mm厚的电极壳体 → 取出电极壳体 → 在电极壳体的下工作面镀紫铜 → 将电极固定座与电极壳体连接 → 构成金属电极。

② 粉末冶金法 用快速成形系统制造母模 → 在母模中充入钨铬钴合金钢(或A6工具钢、铜钨合金)粉用液压机压实合金粉 → 从母模中取出压实后的合金粉模 → 在高温炉中烧结合金粉模构成金属电极。

③ 精密铸造法 用快速成形系统制造母模 → 在母模中充入蜡,构成蜡模 → 用失蜡铸造工艺构成紫铜电极。

④ 浇注法 用快速成形系统制造母模 → 在母模中充入石墨粉与粘结剂的混合物 → 固化石墨粉,构成石墨电极。

⑤ 研磨法 用快速成形系统制造母模 → 在母模中充入环氧树脂和碳化硅粉的混合物,构成研磨模固化研磨模,移去母模 → 在 Hausermann 研磨机上研磨出石墨电极。

7) 用快速成型样件的表面涂覆 为了改善快速成形制件的性能,常常在其表面涂覆一层金属、陶瓷和高分子材料。经过涂覆处理后,能显著提高用快速成形制件的表面机械强度、耐磨性、抛光性,防潮性,工作温度、以及形状和尺寸的稳定性。图 48.2-11 是几种典型的涂覆方法。

① 电化学沉积 (Electrochemical deposition, 简称 ECD) 电化学沉积 (图 48.2-11a) 能在制件的表面涂覆镍、锡、铅、金、银、铂、钯、铬、锌、以及合金,如铅锡合金、NI-SIC、涂覆层厚可达 20~500 $\mu\text{m}$  以上,最高涂覆温度为 60 $^{\circ}\text{C}$ ,沉积效率高。

② 无电化学沉积 (Electroless chemical deposition 简称 CD 又称无电电镀) 无电化学沉积 (图 48.2-11b) 用化学反应形成涂覆层,它能在制件的表面涂覆金、银、铜、锡,以及合金,如 NI/P、NI/P-金刚石、NI/P-PTFE、涂覆层度可达 5~20 $\mu\text{m}$  以上涂覆温度为 60 $^{\circ}\text{C}$  平均沉积率 3~15 $\mu\text{m}/\text{h}$ 。沉积前,表面必须用 60 $^{\circ}\text{C}$ 、pH 值为 12 的碱水清洗 10min,然后用清水漂洗,并用含钯 (PDCL<sub>2</sub>) 的电解液或胶体 (60 $^{\circ}\text{C}$ ) 催化不导电的涂覆表面 10min。

③ 物理蒸发沉积 (Physical Vapor Deposition, 简称 PVD) 物理蒸发沉积 (图 48.2-11c) 在一真空

室内进行,它分为:①热蒸发(低粒子能量);②溅射(中等粒子能量);③电弧蒸发(高粒子能量),包括阴极电弧蒸发和阳极电弧蒸发,典型涂覆层厚为 1~5 $\mu\text{m}$ ,对于最高涂覆温度为 130 $^{\circ}\text{C}$  的阴极电弧蒸发,能在制件的表面涂覆硝酸铬 (CrN) 等材料。通常涂覆层厚为 1 $\mu\text{m}$ ,涂覆前表面须进行等离子体(如 CF<sub>4</sub>/O<sub>2</sub>) 浸蚀预处理 (5min),以便提高涂覆时的粘合力对于最高涂覆温度为 80 $^{\circ}\text{C}$  的阴极电弧蒸发,能在制件的表面涂覆硝酸钛 (TiN) 等材料,通常涂覆层厚为 1 $\mu\text{m}$ ,涂覆前表面须进行等离子体(如 CF<sub>4</sub>/O<sub>2</sub>) 浸蚀预处理 (10min),对于最高涂覆温度为 80 $^{\circ}\text{C}$  的阳极电弧蒸发,能在制件的表面涂覆铜等材料,通常涂覆层厚为 1 $\mu\text{m}$ ,涂覆前表面须进行等离子体(如 CF<sub>4</sub>/O<sub>2</sub>) 浸蚀预处理 (2min)。粒子能量愈高,涂覆时的粘合力愈好,但要求涂覆的表面愈高。

④ 电化学沉积和物理蒸发沉积(或无电化学沉积)的综合 它综合了无电化学沉积和物理蒸发沉积的优点,扩大了涂覆材料的范围(图 48.2-11)。

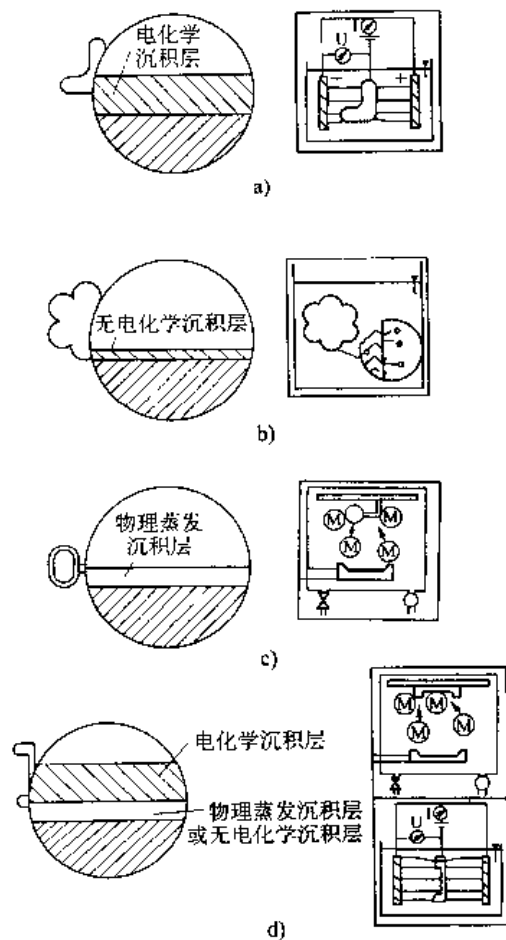


图 48.2-11 几种典型的涂覆方法

## 5 快速成形相关软件技术

### (1) 三维模型表示方法



随着 CAD 技术的发展,出现了许多种三维模型的表示方法,其中常见的有:

1) 构造型立体几何表达法 (Constructive Solid Geometry, 简称 CSG 法) CSG 表示法是运用布尔运算(并、交、差)方法,将一些简单的几何体(如立方体、圆柱体、圆锥体、环等)进行组合,构成复杂的三维模型实体。这种方法的优点是,易于控制存储的信息量,所得到的实体真实有效,并且能方便地修改它的形状,但可用于产生和修改实体的算法有限,构成的图形的计算量很大,比较费时。

2) 边界表达法 (Boundary Representation, 简称 B-rep) 边界表达法是根据几何元素顶点、边和面构成的表面精确描述三维模型实体,这种方法的优点是,能快速地绘制立体或线框模型,此方法的缺点是,它的数据是以表格形式出现的,占用的存储空间大,修改设计不如 CSG 法简单,例如,要修改实心立方体上的一个简单孔的尺寸,必须先用一个实心的圆柱填充这个孔,然后才能绘制一个新孔,所得到的实体不一定总是真实有效,可能出现错误的孔洞和颠倒现象。

3) 参数表达法 (Parametric Representation) 对于自由曲面,难于用传统的几何基元来进行描述,需用参数表示方法,这种方法借助于参量化样条、贝塞尔曲线和 B 样条曲线来描述自由曲面,它的每一个 X、Y、Z 坐标都用参量化形式。各种参量表达格式的差别仅在于对曲线的控制水平,即局部修改曲线而不影响临近部分的能力和建立几何体模型的能力。其中较好的一种是非均匀有理 B 样条法,它能表达复杂的自由曲面,允许局部修改曲率,能准确地描述几何基元。通常为了综合以上方法的优点,目前教有影响的 CAD 系统常采用 CSG、B-rep 和参量的组合表达法。

#### (2) 常用文件格式

快速成形技术中常用的文件格式有:

1) STL (Stereo Lithography interface specification) 格式 STL 格式最初出现于 1989 年美国 3D Systems 公司生产的 SLA 快速成形系统。它是目前快速成形系统中最常用的一种文件格式,用于将三维模型近似成小三角形平面的组合。这种格式有 ASCII 码和二进制码两种输出形式,二进制码输出形式所占的文件空间比 ASCII 码输出形式的小得多(一般为 1/6),但,ASCII 码输出形式可以阅读和直接检查。

2) IGES (International Graphics Exchange Standard) 格式 IGES 是大多数 CAD 系统采用的一种国际图形转换标准,用于支持不同 CAD 系统之间的数据交换。但由于不同 CAD 系统的开发商可能对 IGES 标准的解释不同,从而使各自产生的 IGES 文件有所差

别。

3) STEP (Standard for The Exchange of Product) 格式 STEP 是一种正在逐步标准化的产品数据交换标准。目前典型的 CAD 系统都能输出 STEP 格式文件,有些快速成形技术的研究工作者正试图借助 STEP 格式,不经 STL 格式的转化,直接对三维 CAD 模型进行切片处理,以提高快速成形件的精度。

4) HPGL (HP Graphics Language) 格式 HPGL 是 HP 公司开发的一种控制绘图机运动的语言格式,它已被广泛地接受,成为一种事实上的标准,这种表达格式的基本构成是描述图形的矢量,用 X 和 Y 坐标来表示矢量的起点和终点,以及绘图笔相应的抬起和放下。一些快速成形系统用 HPGL 来驱动成形系统的移动头。

除上述常用的文件格式外,还有如下几种格式:

1) SLC 格式 这种格式是 3D Systems 公司自己的一种文件格式,它是 CAD 模型的 2.5 维表达,能由实体造型软件、表面造型软件或 CT 扫描数据产生。

2) SLI (SLI.ee) 格式 它是一种针对 SLA 快速成形系统的用于控制激光束运动的格式,它由一系列根据切片软件产生的矢量指令组成。

3) CLI (Common Layer Interface) 格式 这是欧洲快速成形行动组织 (European Action on Rapid Prototyping) 支持的一种格式,它试图克服 STL 格式的局限性,希望能为各种快速成形系统所用。

4) CFL (Cubital's Facet List) 格式 这是公司所采用的格式。

#### (3) 常用的 CAD 系统

在快速成形行业中常用的 CAD 系统如表 48-2-1 所示。

表 48-2-1 快速成形系统行业中常用的 CAD 软件

软件名称	开发公司
Pro/ENGINEER	Parametric Technology Co.
I-DEAS	Structural Dynamics Research Co.
UNIGRAPHICS,	EDS-Unigraphics
Solidege	
CATIA	IBM Co.
AutoCAD	Autodesk Co
CADKEY	CADKEY
EUCLID	Matra Datavision
Intergraph	Intergraph Co.
SolidWorks	SolidWorks Co.

## 6 主要的快速成形产品介绍

主要的商品化快速成形系统如表 48-2-2 所示。(摘自 1998 年 1 月美国 CAD/CAM 发布公司、新加坡精技集团有限公司和国内产品提供商宣传材料)。

表 48.2-2 快速成形系统的制造公司和特性参数

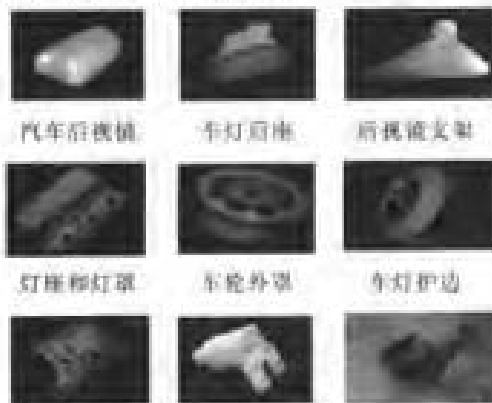
制造公司	型号	上市年	成形方法	所用材料	最大成形尺寸
3D Systems (美国)	SLA-250	1988	液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	250×250×250
	SLA-500	1989			500×500×585
	SLA-3500	1997			350×350×400
	SLA-5000	1997			500×500×585
3D Systems (美国)	Actua 2100	1996	热敏性材料多喷头选 择性喷洒 (MJM)	热敏性材料	250×200×200
Aaroflex (美国)	SolidImager 12	1998	液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	300×300×300
	SolidImager 22	1998			550×550×550
	SolidImager 35	1998			890×550×550
Stratasys (美国)	FDM-1650	1993	丝状材料熔积成型 (FDM)	工程塑料 (ABS)	254×254×254
	FDM-2000	1997		石蜡	254×254×254
	FDM-8000	1997		医用工程塑料	457×457×609
	MedModeler Genisys	1996		专用热敏塑料	254×254×254 203×203×203
Helisys (美国)	LOM-1015	1992	薄形材料 选择性切割	纸质卷材、 薄片塑料	380×250×350
	LOM-203H	1996		815×550×500	
DTM (美国)	Sinterstation 2000	1992	粉末材料 选择性烧结 (SLS)	塑料粉、金属 基/陶瓷基粉	300×380
	Sinterstation 2500	1996			380×330×430
Sanders Prototype (美国)	MM-6Pro	1994	热敏性材料 选择性喷洒 (Ink jet)	热敏性材料	150×150×150
	Model Maker II	1997			300×150×230
MIT-Z Corporation (美国)	Z402	1997	粉末材料 选择性烧结 (SLS)	塑料粉、金属 基/陶瓷基粉	200×250×200
Cubital (以色列、美国)	Solider 4600	1994	液态光敏树脂 液态光敏树脂 固基 (填蜡) 选择性固化 (SLA)	液态光敏树脂	350×350×350
	Solider 5600	1991			500×350×500
SONY/D-MEC (日本)	SCS-1000		液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	300×300×270
	JSC-2000				500×500×500
	JSC-3000				1000×800×500
ESO (德国)	STEREOS DESKTOP		液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	250×250×250
	STEREOS MAX-400				400×400×400
	STEREOS MAX-600		600×600×600		
	EOSINT M-250		粉末材料 选择性烧结 (SLS)	塑料粉、金属 基/陶瓷基粉	250×250×150 340×340×590 720×380×380
CMET (日本)	SOUP-250GH		液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	250×250×250
	SOUP-400				400×400×400
	SOUP-600				600×600×600
	SOUP-850PA				600×850×500
	SOUP-1000				1000×800×500
清华大学 (中国)	SSM-1600	1998	薄形材料 选择性切割 LOM+FDM两功能 丝状材料熔积成型 (FDM)	纸质卷材 ABS丝状材料	1600×700×550
	SSM-600				600×400×500
	M-RPMS-1				600×480×500*
	MEM-250				500×400×450**
	MEM-600				250×250×250 600×400×450

(续)

制造公司	型号	上市年	成形方法	所用材料	最大成形尺寸
西安交通大学 (中国)	LPS-600A LPS-350 CPS-250A	1997	液态光敏树脂 选择性固化 (SLA)	液态光敏树脂	500×400×500 350×350×400 220×220×400
华中科技大学 (中国)	HRPS-1 HRPS-1 HRPS-1A HRPS-1 HRPS-1	1994	薄片材料 选择性切割	纸质卷材	300×300×300 720×400×380 450×350×350 500×400×500 1000×600×500
北京隆源公司 (中国)	AFS-320QZ AFS-300	1994	粉末材料 选择性烧结 (SLS)	塑料粉、金属 基/陶瓷基粉	320×320×425 Φ300×400

## 7 快速成形技术的应用实例

快速成形技术的应用涉及机械制造的各个领域,如汽车、家电、通信、模具制造等行业,在医疗手术方案的制订、生物组织制造也已有很多应用。图 48-2-12 所示为部分用快速成形做的零件。



后视镜内衬 汽车电子零件护罩 牙床的原型

图 48-2-12 部分快速成形件实例

## 8 快速成形技术的发展方向

快速成形技术是近十几年发展起来的一项高新技术,它一方面正迅速在愈来愈多的领域、行业中推广应用,同时本身也在进一步地完善、发展,体现在如下几个方面。

### (1) 新型成形方法的研究和开发

新的成形方法的关键是寻找新的分层叠加成形原理,用什么材料进行分层叠加,如何进行分层叠加是研究的热点所在。因此,除了上述常见的 SLA、LOM、SLS、FDM 和 TDP 等分层叠加成形之外,正在研究新的分层叠加成形法,其目的是进一步改善制件的性能,提高成形精度和成形效率。

金属直接成形,是目前快速成形技术研究的重要方向,这一方向的研究体现在如下几个方面。

1) 大功率激光同轴粉末烧结 美国 Sandia 国家实验室等在能源部支持下,发展 LENS (Laser Engineered Net Shaping) 的快速制造方法,这种方法基于离散/堆积原理,成形的不锈钢、镍基高温合金、H13 工具钢、Ti-6Al-4V 钛合金零件致密度近乎 100%。这一方法正由功能梯度材料制备方面独特的优势的 Optomec Design Company 从事商业开发。

美国 Los Alamos 国家实验室同样在能源部支持下,发展 DLF (Directed Light Fabrication) 的快速制造方法。这一方法用于铝、铁、镍、钛、钽、钨等难熔金属,不锈钢、钛合金等材料的直接成形,开发了一个 1kW 的 Nd:YAG 激光器同轴送粉系统,现由 Synthe Met 从事商业开发。

美国 ArcoMet Corp 在美国防部和海军支持下,发展 14kWCO<sub>2</sub> 激光器同轴快速送粉系统,开发 Ti-5Al-2.5Sn 和 Ti-6Al-4V 柔性快速制造。

Stanford 大学在国防部支持下,正在研制 SDM 方法 (Shape Deposition Manufacturing),采用同轴送粉、高度铣削、支撑铸型,完成了中空叶片、内流道模具 TIS、Mx 高速工具钢的快速制造。

德国 EOS 和美国 DTM 采用 SLS 工艺开发低熔点金属型,美 MIT 采用 3D-P 开发低熔点金属型。

2) 三维焊接 (Shape Welding) 传统堆焊无法精确控制金属微滴之大小、熔深、飞溅、热传导。(送丝速度和焊接电流),Kovacevic 通过调整变压器之频率实现了上述两参数的精确控制。

3) UDS 法 (UDS—Uniform Droplet Spray) 控制金属喷射时的振荡率频率来控制微滴尺寸,电场中控制飞行轨迹。

### (2) 高性能材料的直接成形的实用化

大多数快速成形系统直接成形的材料仅为树脂、塑料和纸等,这些材料与产品采用的真实材料有差距,因此需要进行第二次复制。近年来,出现了象 SLS 这类可直接成形金属件或陶瓷件的快速成形系统,其他

快速成形系统也在寻求直接成形高性能材料的方法,使成形的制件与真实产品完全相符。如涂覆金属、陶瓷的卷材的分层切割叠加成形已经问世,并正在实用化。然而,为了解决成形过程中的粘结问题和减少翘曲变形,所用的金属、陶瓷材料大多是金属、陶瓷与高分子材料(粘结剂)的混合物,因此,成形后须进一步烧结固化,烧掉粘结剂,并在孔隙中渗入金属(如铜)。后处理工序比较麻烦,翘曲变形仍难于控制。

### (3) 成形精度的再提高

目前,快速成形件的精度一般处于 $\pm 0.2\text{mm}$ 的水平。 $Z$ 高度方向的精度更是如此,因为它完全取决于叠层的厚度。影响成形件精度的因素主要有两方面:①由CAD模型转换成STL格式文件,以及随后的切片处理所产生的误差;②成形过程中制件的翘曲变形,以及成形后由于制件吸入水份、温度和内应力变化等不稳定因素所造成的无法精确预计的变形。

对上述第一种情况,正在研制直接切片软件和自适应切片软件,直接切片是不将CAD模型转换成STL格式文件,直接对CAD模型进行切片处理,得到模型的各截面轮廓,从而减少三角形近似化带来的误差。自适应切片是,快速成形系统能根据成形零件表面的曲率和斜率来自动调节切片的厚度,从而达到高质量的光滑表面。

对第二种情况,正在研究、开发新型成形材料,以及成形件的表面处理方法,使成形过程中制件的翘曲变形小,成形后能长期稳定不变形。

### (4) 应用范围的扩大和深入

由于快速成形的巨大吸引力,近几年来,不仅工业界对其十分重视,而且其他许多行业也都逐渐致力于它的应用和推广,因此,应用范围在不断扩大和深入。其中医疗领域里的快速成形技术应用已经异军突起,这几年得到了飞速的发展。快速模具制造行业的应用也是一个具有十分看好的领域,各种新的快速模具制

造方法随快速成形技术发展而迅速发展。可以预料,快速成形技术迅速渗透到更多的领域,将获得更令人鼓舞的应用成果。

## 参考文献

- 1 RP技术与快速模具制造,西安:陕西科学技术出版社,1998
- 2 齐天鹏.自由曲面反求中精度问题的研究:[硕士学位论文].2000
- 3 王军杰.光固化法快速成形中零件支撑及制作方向的研究:[博士学位论文].1997
- 4 王军杰.自由曲面反求关键技术:[博士后出站报告].1999
- 5 胡鑫.反求工程关键技术及其在髋关节生物形态分析中的应用研究:[博士学位论文]2003
- 6 Besl P J, Jain R C. Segmentation through variable order surface fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10 (2): 167~192
- 7 颜永年.全国第二届快速成形与快速制造学术年会报告.2000
- 8 新加坡精技(集团)有限公司.快速成形技术与ZIPPY快速成形系统.1996
- 9 华中理工大学快速制造中心宣传材料.快速成形系统与快速成形/快速制模技术
- 10 宝力机械公司产品介绍资料
- 11 法国MDTV公司北京办事处提供的产品宣传资料
- 12 Starasys公司FDM产品宣传资料
- 13 北京隆源自动成形系统有限公司.AFS快速成形技术与应用
- 14 陕西省激光快速成形与模具制造.激光快速成形技术与设备简介

# 第 49 篇 互联网上的合作设计

主 编 谢友柏  
编写人 谢友柏  
毛军红  
朱爱斌

# 第 1 章 互联网上合作设计的意义

## 1 现代设计一般过程的描述

在网络上进行合作设计,通常被认为是现代设计最前沿的一种设想。所谓现代设计,并不是专指某一类被人们统称为现代设计技术的技术。现代设计首先是指一系列符合时代发展需要的设计观念。当然在这些观念的推动下,无疑会产生许多新的方法和技术,而这些方法和技术本身也是在不断变化和发展着的,但任何一组方法和技术的集合,都不能确切反映现代设计这个词的全部内涵。设计也不限于产品,可以是设计一个产品,可以是设计一个过程,也可以是设计一个机构(组织),它们的基本原理是相通的。

现代设计之所以不同于传统的设计,是由于市场、竞争和技术进步形势的变化,它比过去任何时候都更加依赖于新知识的获取,而不是依赖于经验。各种类型的计算机辅助工具(CAX)的发展,也许使人认为现代设计的特征是越来越多地利用计算机。实际上,这只是一种现象。现在,任何一个技术领域,都是越来越多地利用计算机,或与计算机技术相结合。研究一个领域的发展,是要研究其现在和过去差异的特殊性,而不能满足于把共同的趋势来代替各个领域特点的研究。如果一味到计算机中去寻求解决现代设计中问题的方法,结果就会形成一种误导。设计并不排斥计算,在现代设计过程中,计算所占的比重越来越大。但是设计并不是计算,它不可能由算法上的进展来解决所有问题,这是很明显的。正如,至少在可见的将来,计算机还不能完成代替人的思维。另一点要说的是,任何设计总是从需求出发,而不是从几何(或图形,更确切地说应当称为结构,以下凡提到几何都理解为结构,反之亦然)出发。图是设计的结果,而不是出发点。现有CAD软件的根本弱点是以图为主体的,从一开始就画图。图表现零件和组件的形状,包括三维和动态的图像。这种图形和图像是进行工程分析和视觉感受测试的基础,前者可以通过计算获得零、部件多方面的性能(有人称之为CAE,更确切地说应当是计算机辅助分析),后者则由布置、色彩、光线、动感追求合用及视觉上的美。图还有助于处理零件后续的制造、装配、使用和维护,所以在制造厂中得到广泛的应用。但是它们不能说明为什么要做成这样的形状而不是那样的形状,用这种材料而不是那种材料,采取这些工艺手段而不采取那些工艺手段。CAE是一个单向的过程,它们不能由

对零件和部件,特别是整机的需求产生图形和其他不可缺少的特征参数,因此不能与设计的起点,也就是需求,直接连接。为了说明这些问题,特别是讨论网络环境下设计的合作,首先要分析一下设计过程的各个阶段和各个环节以及其中要解决的问题。

现代产品设计过程可以由以下公式描述:

需求(含潜在的需求)的确认→技术可能(含联想的可能解)扫描→矛盾统一设想(概念)的产生→经济、技术分析(贯穿全过程)→设想的优选和确认→结构的优选和确认→材料的优选和确认→加工过程的优选和确认。

### 1.1 需求的确认

开发一个产品的起点是市场信息分析。当瞄准一个目标市场时,不管是自己的传统市场还是过去没有涉足的新市场,首先要研究已有产品(即竞争对手)在满足用户需求方面存在什么问题,并确定自己的竞争策略,即准备从哪些方面去竞争取胜。构成产品竞争力的要素是多方面的,一般地讲,有以下几个方面:

功能、质量(功能的实现及保持)、价格(全成本、效益)、交货期、售后服务(维修、升级、培训)、环境(含人、机)相容性、营销活动。

通常来说,产品的性能,实际上是指产品的功能和质量两个方面。功能是竞争力的首要要素。用户购买某个产品,首先是购买它的功能,也就是实现它所需要的某种行为的能力。质量则是指产品能实现其功能的程度和在使用期内功能的保持性。虽然上述诸要素的任何一个或几个都能在提高竞争力方面大显身手,但是对于一个企业来说,起根本和持久作用的还是产品的性能。创新是设计的灵魂,首先是指功能上的创新。质量这个概念,在制造业中有各种不同的用法,实际上并没有一个技术上严格的定义。我们把它定义为“实现功能的程度和持久性的度量”,使它在设计中便于参数化和赋值。从设计的观点看,质量是从属于功能的,没有功能也就谈不上什么质量。

在竞争策略上,虽然可以有多种选择,但从根本和长远的角度考虑,还是要生产具有别人不能生产的性能的产品。分析市场信息的结果导向产生一个或几个未来能在竞争中赖以取胜的产品性能作为开发的可选目标,这就是市场需求知识的获取。关于市场需求的知识是设计的出发点,是非常重要的第一步。市场瞬息万

变,关于市场的知识必须随时更新。

## 1.2 技术可能扫描

有了需求,但在技术上和经济上不一定可行,这就是第二步要解决的问题。首先是技术上是否能实现?因为新产品通常不是全部都是前所未有的,许多功能仍可以由过去用过的结构实现。所以设计初始总是在已有知识集中搜索可能解,这就是对已有知识进行扫描,包括在可以进入的知识库中搜索在自己的经验集中搜索。

## 1.3 概念设计

正因为求解的性能是市场上现有产品所不具备的,找到完全满意的解的可能性几乎不存在。于是需要对不能由已有知识解决的部分结构寻求新的解,从而开始了创新的过程。这个过程包括设想采用过去没有用过的新原理、新技术、新结构、新材料、新工艺。这些解决矛盾的设想称为概念,设计的这个阶段称为概念设计。概念设计可能提出几种方案。

## 1.4 技术经济分析

接下来就是对各种方案进行技术上和经济上可行性的测试。所谓技术上的测试,是指用必要的方法去检验它们是否能实现求解的功能和质量,以及实现各种方案的技术和经济的评估,此时往往涉及一系列的CAE分析(为简化文字,认为包括各种视觉和感官以及与环境是否相容的测试,下同)和试验过程。检验并不限于创新的部分,因为局部的创新,常常破坏整体的或其他的功能和质量。测试的方法,有虚拟现实或数字仿真、物理模型试验以及样机试验。不过在概念设计阶段,用得最多的是虚拟现实或数字仿真,当有成熟软件时,它的周期和耗费最少。这个过程明显是一个知识获取的过程。在测试之前,已有知识集中并不存在是否可行的相关知识,而在测试以后,测试结果(确认或否定)就是关于这个问题的新知识。经济上的测试,同样不可缺少。如果所用的原理、技术、材料、工艺成本太高,则也是不可取的。关于某种设想方案成本高低的知识,也是新知识,所以这也是一个知识获取的过程。此外,市场并不仅仅是产品的出口,它还是构成产品的技术、材料、工艺甚至零部件的来源,这些都影响成本的组成。因此,在产品设计的进程中,还要获取市场供应的知识。

## 1.5 详细设计

概念设计不可能涉及全部结构的细节,而是必须集中精力于主要功能和结构的主要组成部分,特别是与创新概念有关的方面。因为方案没有选定之前,繁琐

的细节设计可能会成为徒劳,不同方案的细节设计往往完全不同。对一个选定的方案进行结构细节设计,称为详细设计。当然概念设计与详细设计并不是在时间上完全分割的两大块,更确切地说,它们是前后相继的两个步骤。在局部结构的详细设计中,遇到不能采用已知解时,同样需要创新,这时再次进入概念设计。两个步骤交替进行,直至所有在设计过程中需要确定的问题全部确定为止。这里有三个问题需要说明一下:第一,如果把设计过程看成是一个树形的层次结构,当在低层次的设计中,如果测试失败,就要返回到上一个层次去修改原来的设计,称为回溯。如果无法修改,还要回溯到更上一个层次。第二,当存在可调节参数时,应由优化算法寻求最优解,但要注意局部最优不一定全局最优。第三,作为全寿命周期设计,在设计时要考虑结构、制造、安装、运行、维修(含产品升级)、报废(废品处理和部分再循环使用)等每一个阶段中可能影响功能、质量、人机环境之间关系以及成本等的因素。

很明显,CAE是一个单向的过程,而设计则是不断的求解(产生概念)—测试—回溯—修改—优化—确认的反复过程。虽然CAE以及各种优化方法中包含大量计算,但设计不等于计算。设计需要在广阔的范围中联想(灵感);它同时又是一个知识获取的过程,这甚至涉及到社会活动(例如从用户那里收集产品在运行现场的表现)。二者都不可能仅仅按某种程式运算或推理获得解。图形和图像是设计的结果,而不是它的出发点,设计的起点是对产品性能的需求。

图不是设计的唯一结果,同时要有一份设计说明书,描述图所不能表达的内容。几何特征不是产品唯一特征,虽然在结构设计方面已经获得了广泛的研究和应用,甚至曾经希望它能成为控制整个设计过程的基本特征。但从前面的讨论中可以看出,这显然是不可能的。作为控制整个设计过程的基本特征,只可能是产品的功能和质量及其是否得到了满足,从而提出在现代设计中研究性能特征的需要。

这里没有专门讨论约束问题,但约束是存在的,它通过测试发生作用。

## 2 现代设计的基本特征

从上述关于设计过程的一般描述中,可以得到如下结论:

### (1) 现代设计是需求驱动的

所谓需求,首先是对所设计对象的性能的需求。而驱动,也不仅仅理解为设计任务书上规定的对产品的最终要求,实际上,从概念设计向详细设计发展的过程中,每一个阶段、每一个步骤都是由构成最终要求的各个层次上的子需求驱动的。这一原则,对于以后设计网

上合作者的搜索策略非常重要。

#### (2) 现代设计是以创新为灵魂

创新的目标是要生产出别人所不能生产的产品。创新强调要表现在具有现有产品所不具备的功能和质量上。设计的成败应当由产品在市场上竞争的成败来评价。虽然不能反过来说,竞争的成败就是设计的好坏,但是从列举的竞争力诸要素看,绝大多数要素都是由设计阶段的行为决定的。这就是为什么要在网上进行设计的合作。

#### (3) 现代设计是以知识为基础,以知识获取为中心

现代设计中最活跃的部分是新知识的获取。因为功能和质量上创新的需要,设计知识是一个动态的集合,知识获取是一个要不断进行的工作。对于一个企业来说,设计知识的获取能力是一种综合实力,既包括经营管理,也包括技术水平;既包括资本实力,也包括人才实力;既包括先进的技术装备,也包括长期研究开发的经验。它们的总和,可以看成是一种资源。发达国家的企业多以所谓研究开发(R&D)中心的组织形式来建设和发展这种资源。由于各方面原因,一些企业内部研究开发力量未能得到充分发展,不少企业到目前为止还根本没有研究开发力量。短时期内,要形成这种力量,无论是资金投入、人才集聚、设备建设还是经验积累都很困难。特别要看到,产品竞争的后面,实际上存在着这种更深层次的知识获取资源建设和发展的竞争。企业如果没有一种新的思路,仍旧沿着人家走过的老路追赶,那是很难在竞争中取胜的。阐明这一点,对于了解实现网上合作设计的中心任务、面对的问题和解决问题的方向具有重要意义。

#### (4) 现代设计是对产品全生命周期的设计

这是另外一个需要特别说明的与质量有关的特征:在生产过程和使用过程中,包括报废以后的处理都要与人和环境协调、友好。因此设计时不仅要获取生产这种产品的知识,还要获取所设计的产品使用过程中性能变化的知识、与有关的人的相互作用的知识、与环境相互作用的知识以及报废以后处理的知识。不能满足这些要求,就不是一个符合时代需要的设计,因而也就很难期望其在竞争中取胜。全生命周期的设计使得设计的对象成为一个时变系统,从而使设计的知识获取变得非常复杂。

归根到底,现代设计的特征,离不开知识和知识获取的活动,所以组织网上合作设计,其核心也就是在网上组织知识和知识获取来支持设计。

### 3 设计为什么要在网上合作

讲到合作,对于稍微复杂一点的产品来说,从来都

是合作进行设计的。合作设计并不是新事,过去在计划经济体制下,很多产品和工程都是合作设计的。但是在没有网络传递信息条件的时候,需要把参与合作的人集中到一起工作,或者利用邮政、电报、电话来传递信息。这种低效率的合作只有在低效率的生产中才有勉强存在的可能。因此在大多数情况下,企业总是尽可能地把一切资源都组织在自己的围墙里。这样做的弊病是众所周知的,不必在这里讨论。问题是现在产品更新的速度与那时相比,已不可同日而语,从而对设计效率的要求也日新月异。网络技术不仅为合作设计的快速传递信息提供了可能,而且在得到快速传递信息技术支持的合作设计中,显示了一种大大缩短设计周期、提高设计效率的前景。因为分布式的知识获取资源如果能够借助网络技术通过合作来支持产品设计,那么当一个产品设计完成而需要进入另一个产品设计时,用设计知识获取资源的重组来代替改造,需要的时间和投入都会少得多。

在网上进行合作设计,与必须在知识资源和知识获取资源方面进行合作有关。我国存在着相当一批可以支持在设计产品中获取新设计知识的资源或潜在资源。与工业发达国家的企业不同,这些资源大多存在于科研院所、大专院校、国家和部门的重点实验室、工程研究中心等机构中。拥有这些资源的单位往往是在某些技术领域雄踞一方。它们的总和被认为是一种优势,即“具有比较完整的科学研究体系和雄厚的科技力量”。在国家长期的联系实际、联系生产的方针指导下,与国外专门从事基础性研究的同行之间有很大不同,或多或少都从事过与产品开发有关的工作。如果能把这部分力量有效利用起来,也许能走出一条适合我国国情的发展道路。问题是这些资源目前的状态是分散的,不在企业内部,而且其中的大部分还并没有准备好为这一目标服务,只能算是“潜在”的资源。它们习惯的工作方式比较适合基础性研究,不适合支持激烈竞争中的产品开发。

为使这些不在企业内部的资源能发挥企业内部研究开发中心的作用,特别是对于潜在资源,在互联网上进行合作设计可以很好地解决以下几方面的问题:

1) 能够提供知识服务或/和知识获取服务的资源,以下统称智力资源,其任务是生产“知识产品”,并向企业提供知识服务。它可以作为生产的独立要素存在和发展,为缺乏智力资源的企业服务,向企业提供开发产品所需要的设计知识和获取新知识的能力。

2) 持续地支持这些智力资源的发展。目的是让潜在的能力变成现实的能力,这里既包括采用必要的现代技术措施,也包括改造与此相关的部分管理机制,使它们能够快速响应产品设计任务所提出的需要,特别



是要做好前面说过的已有知识管理工作。因为这些资源是在企业的外部,对其响应速度的要求比对在企业内部的研究开发中心更为重要。

3) 从管理上和技术上使它们能起到与企业内研究开发中心相同的作用。它们要能够以企业开发某项产品的行为为中心和企业组织在一起。共同为这个项目的完成进行设计知识获取,特别是新设计知识获取的工作。

4) 合作的形成、维持和解除应当由市场机制而不是行政机制控制。但在合作的过程中,参与合作的各方都必须遵守一定的规则。社会和政府的任务是提出可供参与各方采用的规则建议,制定必须遵守的规则和监督这些规则的执行,以使合作能顺利地进行,参与各方都能真正从合作中受益。

5) 充分利用网络技术,使合作能够快速和完美地进行,以满足激烈的产品开发竞争的需要。既然是不属于同一企业的许多单位的合作,理所当然也会是处于不同的地域,甚至属于不同的国家,毋庸置疑,信息传递就是头等重要的任务。Web 正好为这种合作所需要的信息传递和交换准备了非常好的技术条件。

做成这几方面工作,并非易事,它牵涉到观念、体

制、管理、投资、技术、利益和安全等一系列问题,解决这一系列问题需要有一个过程。为了一步一步地去做,应当有一个指导方针。当前比较可行的方针是:分散建设,网上组项,市场驱动,强强联合。

还有一点需要说明的是,在网络上进行合作设计首先是从我国的实际状况提出的命题。自 1996 年起,西安交通大学润滑理论与轴承研究所开始了这方面的研究工作,1999 年教育部批准成立了现代设计与制造网上合作研究中心,其三个成员单位是西安交通大学、上海交通大学和华中科技大学。该中心的中心(西安)和中心(上海)运行着现代产品设计与研究开发网络([www.chinamoderndesign.com](http://www.chinamoderndesign.com)),很好地示范了基于网络的合作设计,并且已经具备一些产品创新设计能力。值得一提的是,在工业发达国家,有关设计理论与应用的研究越来越多地注意到在网上进行合作设计的优越性,美国国家研究咨询委员会(NRC)和美国国家工程院(NAE),受美国国家宇航局(NASA)委托,在 2003 年完成了一份题为《先进工程环境(Advanced Engineering Environment)》的咨询报告,指出网络技术将给产品设计与开发带来巨大的变革,必须加强有关的研究与应用工作。

## 第 2 章 设计知识服务和分布式智力资源

### 1 设计中的知识

#### 1.1 产生知识的信息源分类

从设计的过程和要达到的目标看,现代设计的特征离不开知识和知识获取的活动,所以组织网上合作设计,其核心也就是在网上组织知识和知识获取来支持设计。因此有必要仔细讨论一下设计所需要的知识、获取这些知识的过程和所需资源的构成。

设计一个产品的知识,通常产生于以下几方面来源的信息:

- 1) 已有知识;
- 2) 市场信息;
- 3) 由虚拟现实对可能解进行测试和评估所得到的信息;
- 4) 由物理模型试验对可能解进行测试和评估所得到的信息;
- 5) 由样机试验对可能解进行测试和评估所得到的信息;
- 6) 已经投入使用的产品的运行表现信息。

#### 1.2 设计知识的结构特征

设计知识是一个动态集合。竞争压力迫使制造业对新创意、新技术更为强烈的追求。昨天是新的,今天就是众所周知的了,明天又要被另一知识所代替。竞争取胜靠的是只有少数人掌握,甚至独有的知识。过去用垄断或专利保护延长少数人掌握或独有的时间,随着知识获取速度的加快,这种延长越来越困难。因而,必须再去获取更新的知识。所谓以知识为基础,指的是这样一种动态存在的知识集合。与之相对的是基于经验的设计,以为有一本手册就可以设计了,这是过去占统治地位的观点。

另一方面,创新是设计的灵魂。有人以为引进一套软件就可以设计了,这是现在很多人迷信或竭力宣扬的观点。实际上,不论是数据库、专家系统、虚拟现实工具等,都是建立在已有知识基础上的。创新时,面对的是未知世界,已有知识仅仅是用以获得理解和预测这个世界的知识的一个信息源,还要融合从其他信息源得到的信息,才能产生创新所需要的新知识。所以设计是以新知识获取为中心。数据库、专家系统、虚拟现实工具等都必须随时用新获取的知识加以更新,虽然在设计完成后,为此而获取的新知识又变成了已有知

识,但是与一个久不更新的知识集合相比,显然一个不断更新的已有知识集合更支持设计竞争。

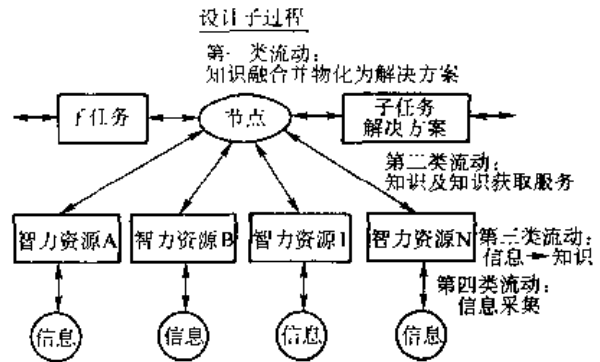


图 49.2-1 设计过程中知识流动的一般路径

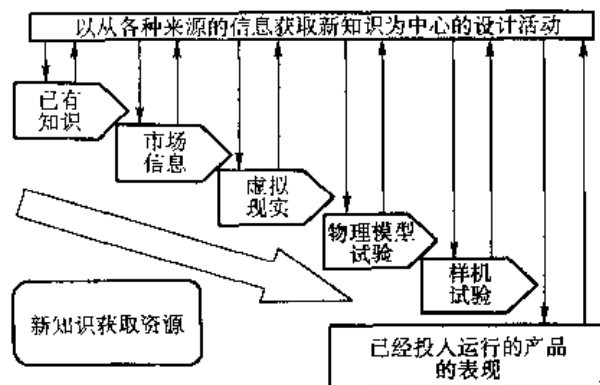


图 49.2-2 知识获取所依赖的各种资源

设计过程中的知识流动可以分成四类。图 49.2-1 给出了从形成设计任务到得到解的知识流动的过程。对于子过程也是这样。由已有知识联想产生概念,由对概念解评估的需要,搜索有关的智力资源单元并请求知识服务和知识获取服务,对各种智力资源单元提供的新知识,即从不同信息源和不同视角得到的评估结论,有一个融合的过程,包括比较、综合、优化、摒弃、回溯和再设计等等,最后物化为解决方案,这是第一类流动。知识和知识获取是资源依赖的,知识和获取的新知识由分布的资源单元汇集到设计过程中,流动包括与发现(搜索)、描述(发布)、测试和形成或解除委托和服务关系等一系列活动内容,这是第二类流动。第三类流动是在各个智力资源单元内部进行的,根据请求方的要求采集信息并加工成可以支持设计的知识,不仅有内容的要求,还有形式的要求。一种是不需要物理过程就可以得到的数字化知识,要以交互可视化形式

提供;另一种是必须经过物理过程才能提供的知识,也要求研究如何更充分地利用信息技术的成果。图 49.2-2 示意地给出了设计工作中用到比较多的信息源以及它们对资源的依赖。第四类流动是各个智力资源单元根据需要采集信息。没有信息,知识就没有来源,这是一个基本原理。图 49.2-2 中箭头用了较粗的线,说明这种流动的不可替代性。

## 2 获取信息的资源

### 2.1 虚拟现实需要的资源

数字仿真或更高级的虚拟现实,都是建立在一系列数学模型(即已有知识)的基础上的。根据给定的系统结构和对系统的输入,预测系统的性能。它是获得关于一种新构想或新概念是否可行的知识的有效工具。因为主要是在计算机上操作,通常不制作专用的模型和实物,在硬件和软件的配置上具有很大柔性,因而能节省时间和资金,可以为设计人员在考察其设计时大规模地运用。这种关于新构想或新设计的知识,和经验相比,就属于新知识范畴。

数字仿真和虚拟现实在知识获取方面具有巨大的潜力,有很大的发展空间,是当前的研究热点。但它绝对不是万能的。如果真正要把数字仿真和虚拟现实当作设计知识获取的一个全面有力的工具,而不是仅仅作为某些狭窄目标知识获取的工具,那就必须面对如下事实:随着对产品性能要求的不断提高和对自然规律认识的不断深化,人们总是处在无数学模型和有数学模型,旧的数学模型和新数学模型不断交替的过程之中。所有新发现的现象或新构想从一开始都没有数学模型或没有准确的数学模型。这里可以说一说“摩擦学设计”。由于一个机械系统的摩擦学性能及行为有强烈的系统依赖性和时间依赖性,同时它们又是分属于许多不同学科研究的过程耦合的结果,所以摩擦学问题的数学建模十分复杂。例如即使是一副简单的试样,在一种系统条件(例如 Timken 试验机)下获得的结果,往往不同于另一系统条件(例如 SRV 试验机)下的结果,当然也不同于待设计的目标系统条件下的结果;另外,对于新系统、跑合系统、磨损系统的结果也不一样。为了仿真的需要,不仅要有系统行为本身的数学模型(这个模型涉及到许多不同学科研究的问题),还要有系统条件转化的模型和时变规律(为全生命周期设计服务)的模型,否则仿真所做的预测就不准确。这个事实一方面告诉我们,在讨论建立一个无所不包的模型,也包括讨论建立在数学模型基础上的各种优化研究时,要持慎重态度;另一方面也为我们提供了现代设计研究几乎是无限的领域,因为产品设计总是要

求提供的设计知识越来越逼近真知,给出的预测越来越精确。

要对大系统和复杂过程进行数字仿真和虚拟现实,不仅仅是数学模型和计算机运算的问题,还涉及到多媒体技术、传感器技术、伺服技术等。某些虚拟现实系统具有非常复杂和庞大的结构,而且是十分昂贵的。有的还带有部分物理模拟的特点,可以说是一种混合模拟。例如模拟宇宙载人舱。当然与发射一个真的载人舱到空间去相比,花费的时间和资金已小得多了。吉林大学汽车动态模拟国家重点实验室(ADSL)的一个汽车动态模拟装置,由一个汽车可以开进去的“模拟仓”(球壳)和下面若干个计算机控制的液压执行器支撑,驾驶员在车内可以看到 145°视角中移动的道路(场景)并操纵行车,计算机则由虚拟路况、操作行为和该车的系统参数确定执行器的动作,推动球壳和车做相应(实在)的响应并伴有音响,可以从包括亲身感觉在内的信息来评估人一车一交通环境闭环系统的整体性能。

数字仿真和虚拟现实虽然比实物试验要节省时间和资金,但是往往仍然非常昂贵。首先,需要基于先进数学模型的软件平台。随着仿真功能的提高,软件的价格越来越高。其次,计算机硬件,也随着仿真对处理速度要求的提高而变得很贵。第三,对于混合模拟,还需要许多其他的硬件装备,如特殊的传感器、特殊的执行器以及各种信息传输的设备。第四,最重要的是能操作这些设备的人和他们对问题的元知识和领域知识的了解以及经验。上述的软、硬件并不是买来就可以用的,复杂的系统一般需要在系统上工作两年或更长时间才能熟练使用,而对专门问题的二次开发,则要在两年之后才能进行。要对模型进行校正,还需要做专门的试验,例如测定某个界面上的传热系数或研究某个力学系统边界上的约束条件,并经过反复的考题和与实际核对后才能使用。这就是受过(专业)教育的人作为载体的重要性的明证。

一般说来,越是通用的资源,其功能和质量都相对较低,但是使用的频数会较高。为了提高功能和质量,需要进行二次开发,投入大量的人力和资金。经过二次开发以后,所开发功能的使用频数无疑将会降低。专用的资源功能和质量十分诱人,不过价格昂贵,使用频数低。随着技术发展,版本升级会十分频繁,维持这种资源的正常运行很不容易。

因此建设和有效运行一个数字仿真和虚拟现实的资源也绝不是轻松的事。无怪乎企业常常用有这样或那样的数字仿真和虚拟现实资源作为竞争的资本,并投入大量资金去发展这种资源的功能和质量。反过来,如果一个企业已经建设了针对某一类产品的这种资源,是否都能充分、有效地使用,通常并不理想,特别

是规模较小的企业。而在改变产品类别时又会成为一种阻力。

数字仿真和虚拟现实是获取新知识的有效手段,而且通常是在测试和评估的一开始就使用,采集到比较多的信息后,才转入其他手段以求核对,但是这个手段靠的是基于已有知识的模型;虽然比较省时省钱,但仍依赖将在后面说明的具有6个不可分要素(人、资金、设备、经验、管理、技术)的资源。

## 2.2 物理模型试验需要的资源

数字仿真和虚拟现实是以模型为基础的,而模型则是已有知识。采用基于已有知识的模型去测试和评估过去没有设计过的产品,其应用范围当然要受到限制。上面说的传热系数和边界条件就是如此。摩擦学行为具有明显的系统依赖性和时间依赖性,当系统条件和时间改变后,模型也必须做相应的修改。而修改的根据就是物理模型试验。例如原来用于设计每小时行走100公里的汽车的数字仿真和虚拟现实平台,在设计每小时行走200公里的车时是否能用?需要通过物理模型试验来确定哪些方面会有变化,并最终确认修改后的软件平台的可用性。所以,重要的新设计的子系统和部件都要经过物理模型试验以获取对可能解测试和评估的信息。

物理模型试验需要设计试验、组织试验和操作试验的人,需要试验台、测试仪器、数据处理设备和试验场地。例如造船工业中的“船池”和航空工业中的“风洞”,都是昂贵的大型物理模型试验的资源。这些资源所需要的投入,与上述同等规模(指被测试对象)的数字仿真和虚拟现实资源相比,一般来说要大得多,建设和变更的时间要长得多,对运行这类资源所需要的人的专业技术种类和水平要高得多,熟练运行这类资源所需要的经验要复杂得多。上节最后一段所论述的问题,在这里表现得更为突出。发展可以为众多资源需求方服务的物理模型试验资源,应当是节约资源和优化资源的方向。

现在有人建议一些远程(物理模型)试验的研究项目。实际上新知识获取中的物理模型试验与教学中的模型试验是有区别的。它是为了获取一个关于未知设计可能解的新知识,因此试验本身需要重新设计,而往往不是用一个已有的试验。这意味着即使有了已经很成熟的通用试验台,仍旧需要设计和制造试件,研究并确定加载方式和边界条件,研究并确定测量参数,选择测量仪器和测点位置,研究和确定数据处理方法以及规划试验程序等。针对每一个不同目的的试验都是不相同的,这种试验离不开试验资源的拥有者和运行者的直接操作。也有一些自动化程度很高的仪器,项目的

建议者们想通过互联网实现对它们的远程控制,由用户自己来操作。这种用户直接操作的价值和对于那些贵重仪器的拥有者是否愿意把自己的仪器放在网上让没有资质证明的人操作还是问题。目前通行的办法是由用户将自己的需求和样品经过互联网传送到资源的拥有者或运行者手上,由他们设计试验和熟练地操作试验,并把结果返回到用户手里。当然,这并不排除用户可以通过摄像头、浏览器和话筒监视试验的进程和与试验的运行者讨论试验的有关问题。

重要的新设计的子系统和部件都要经过物理模型试验以获取对可能解测试和评估的较直接的信息。它的资源建设和运行、维护比虚拟现实和数字仿真更难。

## 2.3 样机试验需要的资源

样机试验需要相应的场地,例如对于汽车制造业的跑车场,因为要模拟各种实际路面,规模也是非常大的。没有跑车场的时候,车设计出来后只能在普通公路上做样机试验,而一些大的汽车制造企业,一个企业就拥有多个跑车场,甚至可以用从目标市场测得的道路谱,在自己的跑车上模拟这种道路,以求产品性能适应目标市场的道路。即使是物理模型试验,模型设计仍旧有假设的部分,特别是边界条件,它们是建立在已有知识的基础上的。因此重要产品都要经过样机在服役条件下试验的阶段以便对可能解获取最终知识。数字仿真和虚拟现实平台,如果不经过物理模型试验和样机试验的核对,就不可能证明它们的价值。虽然有的数字仿真和虚拟现实软件并没有直接进行这项工作,但它的正确性是在长期使用中被用户的实践(物理模型试验和样机试验,甚至包括已经投入使用产品的运行表现)结果所证明。许多数字仿真和虚拟现实软件平台之所以要求用户自己做二次开发,就是需要用户根据自己实践(物理模型试验和样机试验,已经投入使用产品的运行表现)中得到的信息来构造模型或使平台中的模型精确化。

创新设计产品的系统性能,需要样机在服役条件下的试验中测试和评估。样机试验与产品的联系十分密切,一般都具有比较高的专业性,通常在开发该产品的企业内进行。

## 2.4 在产品运行状态监测需要的资源

这个信息来源过去并不为许多制造商重视,虽然希望创名牌的公司经常做用户调查,当发生问题时,立刻派人到现场去解决问题,但是很少把它与为获取设计新知识的采集信息直接联系在一起。任何产品都要求在全生命周期中保证性能。这使得原来把对象看成是时不变系统的理论和方法,都要做应有的发展,这在

目前还是一个巨大的未知空间。前面说到的信息,其中有一类就是关于产品各种性能特征随时间变化的信息,如图 49.2-3 所示。过去的力学性能,是以出厂时检测合格为目标,而现在则需要控制整个生命周期中性能的变化和衰退。现在机械电子产品普及到个人、家庭生活的每一个角落而不是像过去那样仅仅存在于具有专门技术的工厂里,这是一场意义深远的革命性的变化。全生命周期中,产品在用户手里使用的阶段是最重要的阶段。用户购买产品是为了使用产品的性能,使用时的表现,已经成为制造商竞争的焦点。历史上制造商们从来没有对这个阶段给予像现在那样的重视。个人和家庭用户,不可能像工厂用户那样对每天要用到的数十种甚至上百种不同的机械电子产品具有专业知识,大多数情况下可能是一点知识都没有,因此使用和维护成了突出的问题。特别要提到的是对个人和家庭来说通常是没有备份的,如果发生故障停机,就会造成极度不便,零故障停机是当然的目标,少维修、免维修、预测维修和自维修就成为设计的重要命题。维修任务从用户转到制造企业之后,维修的成本也就正式进入了产品的成本。所以制造企业必须尽量减少这一部分的开支,包括从设计时就要加以考虑,这就是为什么全生命周期设计将成为企业的自觉追求,而不再仅仅是用户的愿望。但是,这一类信息随产品而变,随使用的条件而变,十分复杂,而且只能在离散的用户手中,即使用现场以及在全生命周期中跟踪采集,目前也还没有成熟和普遍的技术。

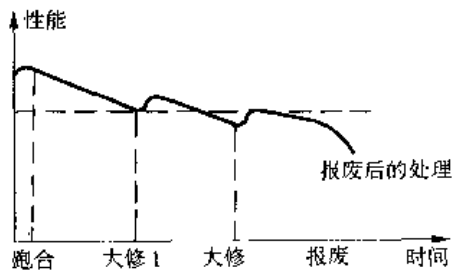


图 49.2-3 全生命周期中产品的性能与时间的关系

现在的状态监测技术还仅仅在非常有限范围中使用,应用的目的是为了防止灾难性事故,而不是为了掌握产品全生命周期中性能变化的信息。状态监测技术还需要从根本上发展,才能满足这个要求。这是一个亟待解决的命题。目前这一类资源还是与产品密切联系在一起,特别是与产品设计密切联系。如何能与产品分离,成为一种通用的技术,也还需要进一步的工作。

从上面的讨论可以得到以下结论:(1)已有知识的使用和新知识的获取都必须以人为本而不是排斥人,计算机以它独具的能力,起支持作用。计算机的工作需要人做前期准备,相当于对人的教育。要求计算机工作的集成度

越高,“教育”的投入越大,因为它离不开受过教育的人来做“教育”工作。从全局考虑,集成度与效益之间有一个最优关系。这是后面将要讨论的分布式资源理论发展的基础。(2)知识最终是以人为载体,但要集成起来才能在设计中起作用,所谓在互联网上合作设计,实际上就是从资源的组织和利用角度来解决集成的问题。新知识获取依赖具有 6 个不可分要素的资源。

## 2.5 其他信息资源

### 2.5.1 已有知识

已有知识即通常所说的“经验”,它应当理解为包括设计上一代产品的全部知识。已有知识当然既包括元知识,也包括领域知识。已有知识只是产生可能解的信息源,因为不是“确解”,所以不能列为知识,但是它们是获取有关新知识和产生设计概念(可能解)的基础。人在一定历史时刻,不可能得到绝对完整、正确的知识。例如协和超音速客机,在运行多年之后,因为一次意外事故而发现设计当中的缺陷,并在改进后才再次起飞,但是最终又加以放弃。也就是说,在现实中,人们只能掌握相对完整、正确的知识。但是,每次获取新知识以后,人们就在知识的完整性、正确性上前进了一步。可能解是由相对完整、正确的已有知识产生的,虽然经过新知识获取过程得到“确解”,最终设计同样只是相对完整、正确的知识。成本和时间不允许追求和等待绝对完整、正确的结论,设计总是在知识的相对完整、正确中做出决策,这就是产品创新设计过程中的风险所在,也是不能排除人,而仅仅研究计算机的理由之一。在产生可能解的过程中,如果知识完整、正确的程度越高,在后面阶段中设计的回溯会越少,设计失败的风险也越小。但是,完整、正确程度越高,意味着满足新的性能需求越少,获取并注入产品的新知识越少,也就是说,创新的程度较小,在竞争中的优势也会较小。这里所讲的已有知识,指的是全人类所具有的知识。而对于一个具体的人,或一个设计团队,他或他们的已有知识则是一个小得多的子集。要降低风险,就不能完全依赖自己的已有知识。所以已有知识是十分重要的。已有知识可能以下述形式存储:

- 1) 数据;
- 2) 文件;
- 3) 图。

还有两种讨论得比较少的特殊形式要在这里做进一步的研究,即

- 4) 已有产品,作为已有知识的载体,其中隐含着已有知识;
- 5) 虽然还没有变成产品,但是已经完成的设计,

作为载体,其中同样隐含着已有知识。

之所以说是隐含,这与知识产权保护有关。和数据、文件、图不同,它们给出的是“显含”的信息,对于已有产品或已有设计,即使拿到了新产品或者新产品的图纸,并不能拿到全部设计知识,即并不能知道所谓的“怎样知道”。加上专利法的规定,也不允许使用受保护的知识。以为买了别人的设计,就得到了全部设计知识,其实不然。

在产生可能解的阶段,并不需要、也不可能将所有部分都重新设计,在后面的阶段中也是一样。在新的性能需求驱动下,通常仅仅是其中一部分需要重新设计,其余部分或者是采用现有的产品或设计,或者对现有的产品或设计进行某种改进。而这种改进,甚至于重新开发,考虑到已有的知识资源,往往也会请求已有产品供应商提供重新开发的服务。

需要特别指出的,即使是已有知识,同样也是一个动态的集合。新知识用过以后,就成为已有知识。新知识不断产生,已有知识的集合也不断扩大。不管用什么形式存储已有知识、维护和更新已有知识的知识库,包括最新的产品目录,是一项极重要的任务。当一个知识库停止了更新,那么其生命就已经终止,最后将在竞争中被淘汰出局。

从前面的实例可以看出,知识获取的能力是一种综合实力,既包括经营管理,也包括技术水平;既包括资本实力,也包括人才实力;既包括先进的设备,也包括过去的经验。归纳起来是6个要素:人、资金、设备、经验、管理和技术。采集、管理、维护和更新已有知识的存储,同样是十分艰难,同样需要这6个要素。许多企业,希望在员工离开时能把知识留下来,做了各种尝试。因为这涉及到深度的元知识和领域知识,不是随便什么人都能准确地把每一条新采集到的知识表达和存储的。由获取这些知识的本人来做,效果也许较好。智力资源有两个实际上存在的特点,不能不予以重视:一是智力说到底是以单个人为载体,却需要集成起来才能发挥作用,所以上述6个要素又具有不可分性;二是智力是一个动态的概念,昨天先进的,今天可能不再先进,明天就落后了,所以说知识是一个动态的集合。由于涉及到多方面的利益特别是作为载体的个人的利益,有效地采集、管理、维护和更新已有知识的存储,仍然是一个十分复杂的问题。特别是涉及到知识产权问题,它越来越受到法律保护。总而言之,关于已有知识的讨论,可以归纳为三点:作用重大、浩如瀚海、有待用好。它本身是一个巨大的资源,而管好、用好同样

需要有相当的资源支持,即需要6个要素。根据这里及将要讨论的理由,已有知识资源同样是以分布式的存在较为合适。

已有产品和已有设计中都隐含着已有知识;但是说到底,智力是以单个人为载体,却需要集成起来才能发挥作用,同时它又是一个动态的集合。

## 2.5.2 市场信息

市场信息包括已有知识和不断采集到的新信息两部分,由于其对产品创新的重要程度和与技术领域的信息性质有所不同,所以专门列出一条。前面多次提到性能需求驱动设计的思想,而这种对需求的认识、搜索满足需求的可能解和搜索能提供知识或获取新知识服务的资源,都需要市场信息。

设计所需的市场信息至少有四方面内容:

1) 需求信息。当然这种信息应当是由经营人员或某种智能系统处理过的,以确定或模糊的形式给出,使设计人员便于应用。

2) 供货信息。前而说过,在新的性能需求驱动下,仅其中一部分需要重新设计,其余部分或者是采用现有的产品或设计,或者对现有的产品或设计进行改进甚至重新开发。考虑到已有知识资源和知识产权问题,也往往是请求已有产品供应商提供服务。不仅制造过程中有合作,产品设计和知识获取方而也将会有更重要的合作,这就需要一种更广泛的市场信息资源。有一项统计表明,制造业的开发活动中有40%~70%依赖外源(外部供应商、合同制造商、合同设计服务公司)。以汽车工业为例,在成本的分布中:

14% 间接成本(厂房、办公等);

20% 直接采购成本(通过叫价竞争,可降低的空间已经不大);

66% 直接设计成本(难以实行叫价竞争,变化的空间很大)。

3) 成本信息。不仅是企业内部为设计、制造、售后服务所需要的成本,还包含其他所有可能合作的过程需要的成本。

4) 竞争信息。其他企业投入或即将投入市场的同类或代替产品的性能、价格等方面的信息。

市场信息在资源和处理上的需要,与其他领域的基本相同,不再重复讨论。

市场需求是设计进程的驱动力,同时也是服务请求方和服务提供方合作的纽带。

### 第 3 章 分布式智力资源的运作模式

#### 1 智力资源的构成——服务提供方

##### 1.1 智力资源的构成要素

前面讨论了可以提供知识服务的智力资源和新知识获取服务的智力资源的内容，它们可以用 6 个不可分要素来概括，即人、资金、设备、经验、管理和技术。之所以说不可分割，是因为缺少其中任何一个要素，资源就不能正常工作。传统上，企业生产自己产品所需要的资源基本上都配置在企业的内部，即采用一种所谓垂直的资源结构（如图 49.3-1 所示）。垂直资源结构在市场需求多变的今天，是必须加以改变的。否则，一个资源还没有建设到可以发挥作用，可能就已经过时了。对于已经存在的资源，为了充分收回投入，提高资金的利用率，企业往往难以下决心改变已经没有竞争力的产品，不能迅速转而开发新的更有竞争力的产品。所以分布式的资源结构（如图 49.3-2 所示）已经越来越受到重视。互联网技术的发展为分布式资源结构的应用创造了前所未有的有利条件。这里定义作为在一个知识领域中，可以提供知识服务的智力资源或知识获取服务的智力资源服务的最小单元，称为资源单元。参与产品创新的资源单元可能属于几个不同的企业，也可能是独立的。下面将要研究作为一个最小资源单元的组成要素和可以独立存在的条件。

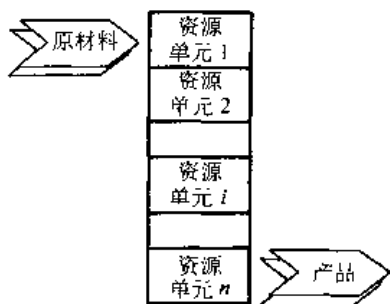


图 49.3-1 企业的垂直资源结构

图 49.3-3 给出一个最小资源单元的组成要素。人是首要的要素，这里不研究离开人的计算机系统，因此没有采用智能体（Agent）这个名称。前面说过，人是知识的载体。当然人在某一个领域中是有知识的，如果有一个以上的人在这个单元中工作，那么他们还必须按照规则组织成一个团队。其次，必须具有可以提供已有知识或获取新知识服务的能力。第三就是通信

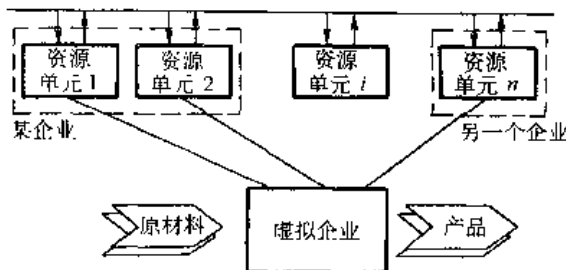
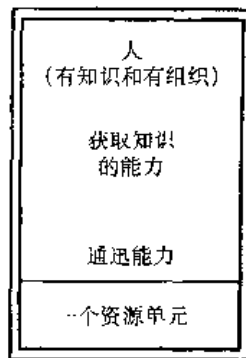


图 49.3-2 企业的水平资源结构



49.3-3 资源单元构成要素

能力。在产品阶段，在各个单元之间运动的主要是信息流而不是物流，虽然并不是没有物流。这就是设计与加工的不同，它可以更大程度地利用互联网。

通信条件视各个单元之间所要传递的信息种类的不同而不同。因为现今商品软件平台的相互封闭，特别是造型软件所产生的信息交流，在技术上仍相当困难。虽然有 STEP 或 VRML，因为各自的定义不同，仍旧难以沟通，有时可以浏览，但一般不能操作。一种可能采取的办法，是用多种软件平台生成的造型为不同的用户服务。例如，在提供服务信息时，让请求方可以从浏览器下载所选结构造型并进入正在设计的装配图以供测试。由于用户装配图可能基于不同的造型平台，如 AutoCAD、Pro/E、UG、Solidworks 等，为了可能对下载的 2D 或 3D 造型进行操作，服务方要准备适合各种平台的结构造型，让用户可以根据自己所使用的平台下载，这种做法对资源单元提出了过分的要求，只能看作是一种过渡办法。对于非结构性信息，则可以转换成 XML 形式加以传递。

##### 1.2 智力资源的生存条件

资源单元（这个资源单元至少应当在技术上

能够提供知识服务)在市场经济下的生存条件,可以说明如下:

- 1) 提供服务的知识必须是最先进的,而且要能够保持其先进性;
- 2) 必须有高质量和高效率的服务以求高额回报;
- 3) 必须有高额的回报以驱动资源的拥有者和运行者极力保持知识的先进性和服务的高质量和高效率;
- 4) 要有充分的服务请求。

## 2 设计实体(服务请求方)的构成要素

服务请求方的最小单元称为设计实体,其结构如图49.3-4所示,更大的单元或实体可以看成是最小单元的集成。

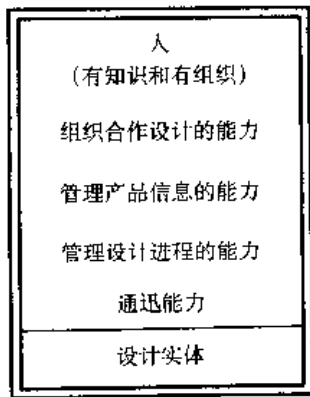
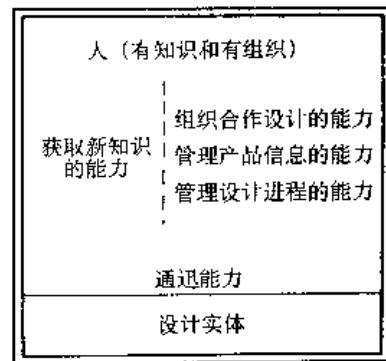


图 49.3-4 设计实体

关于设计实体的特征可以提出如下几点:

- 1) 有一个设计任务;
- 2) 直接从事当前任务的设计或部分设计工作;
- 3) 对任务的完成和所设计产品在市场上竞争的成败负责;
- 4) 需要其他设计实体和/或资源单元的服务来完成设计任务。

当然,设计实体和资源单元也可以存在于一体,但这并不妨碍各自具有自己的特征和在自己特征支配下的行为,如图49.3-5所示。



49.3-5 资源单元与设计实体的组合体

## 3 合作设计的层次结构

在互联网上,分布的服务提供方(资源单元)和服务请求方(设计实体)按图49.3-6所示的层次结构联合工作,现说明如下:

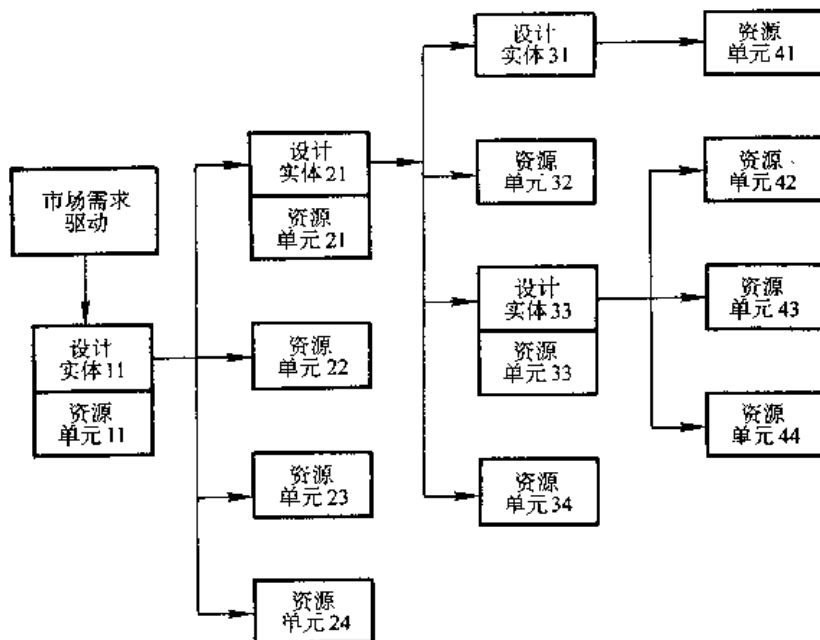


图 49.3-6 合作设计的层次结构



1) 由对市场需求认识而产生的设计欲望并经过利益与风险平衡形成的一个设计任务,可能是一个实体,也可能是几个由合同(规定权利和责任)联系起来的实体,它们是第一层。这里不包括若干相互竞争的实体和任务。

2) 由于第一层的实体不具备使可能解具体化和进行测试、评估、优化和再设计全部过程所需要的资源,从而需要别的资源单元提供服务。于是产生了第二层和其他后续层次的服务提供方和服务请求方。

3) 提供服务(包括提供产品)并不等同于提供隐含的知识,更不涉及有关知识服务智力资源和知识获取服务智力资源的转移,所以不能合并为一个层次。显然,服务提供方也可能是一个设计实体,它提供改进甚至重新开发某个已有产品或已有设计(通常这个产品或设计是该设计实体所拥有的)的服务,而这个产品或设计是正在进行的设计任务的了任务。

4) 关于进程管理,服务请求方与服务提供方由合同规定各方的权利和责任,并按合同进行管理。合同应当包括完成的质量、数量、时间、支付的约定、违约和发生不可预见情况的处理。但是请求方并无权了解和控制提供方的进程和实现的技术细节(除非另有约定),所以也不能和不需要合并为一个层次进行管理。

根据这些认识,可以明确以下一些原则。

1) 动态联盟的组成,说到底是在发出服务请求和响应服务请求并最终确定合作的过程。性能需求驱动的原理适用于合作设计的各个层次。“父层”的设计欲望和设计任务固然由对市场需求的认识和利害平衡产生,“子层”与“孙层”也遵循同样的原理产生自己的设计欲望和设计任务。因为上一层发出的服务请求是对下一层的一种市场需求信息,而下一层则根据这个信息产生自己的设计欲望和设计任务。

2) 上一层在产生设计任务后,首先要在概念设计中明确哪些子任务将由自己完成,哪些将向市场请求(下一层的)服务,这是概念形成过程中一项必不可少的工作。大多数情况下在从设计欲望向设计任务转变的过程中,即在对利益与风险分析、评估并最终取得平衡的过程中就需要基本完成这一项工作。前面说过,产生一个设计任务的过程也是一个设计,它具有产品设计过程的所有特征,这个工作主要还是由人来完成。当然,如果是几个人在不同的实体中或不同地方共同完成,那么目前支持互联网的商品软件可以满足这个需要。

3) 发出的服务请求和能够提供服务的清单可以送到一个公共的网页,现代设计与制造网上合作研究中心(<http://www.chinamoderndesign.com>)推荐一个 UDDI 协议并提供了一个专业的 UDDI 注册中心,具体注册办法在第 4 章中说明。

4) 合作设计的每一层次都是为满足性能(功能和质量)需求提出服务请求和响应服务请求,为了更有效地实现性能需求驱动,便于搜索引擎的设计,建议制定描述产品性能特征的标准并由性能特征作为请求服务方和提供服务方之间的联系。作为一个完整的设计任务,性能当然专属于产品,而在设计过程中,性能特征的描述也适用于非产品服务。

5) 发出服务请求和响应服务请求并最终确定合作的动态联盟并不是一下子就能形成的。对于一个设计,特别是有较大创新的设计任务,涉及到的资源结构与传统垂直资源结构或者是加工过程所具有的特点完全不同,在合作设计开始时并没有一个完整的动态联盟,联盟往往是在确认需求后才请求服务,服务提供后即结束合作。由于回溯而改变合作伙伴关系也可能发生。在一个合作项目中,只有一部分设计实体和资源单元始终存在于联盟中,另一部分则仅仅在联盟中存在一个时期。所以“动态”不仅理解为在从一个产品合作设计到另一个产品的合作设计的变化,更可以理解为在一个产品合作设计的不同时期中的变化。

6) 某些“子层”与“孙层”的联盟可以看成是子联盟,它可以作为一个单元计入上一层的联盟。出于对知识产权的考虑,它的组成和运作对上一层和更上一层往往是不透明的。所以并不需要一个统管全局的动态联盟管理系统。在所有节点上动态联盟的管理软件可以是结构相同而简单,主要只管理自己“子层”中的合作者(不讨论单元内部的管理)。在第 5 章和第 6 章中将介绍有关的管理平台和用法。

7) 合作是从服务请求方提出需求和服务提供方对需求响应开始的,但是双方在技术条件、任务进程、支付条件等方面的认识并不会完全一致,这是一个双向选择的过程,需要经过竞争(招标)和协商取得一致。这也需要时间。如果资源单元对自己所能提供服务的规范和条件有比较明确的描述和服务请求方对自己的需求概念清晰,并且双方对于“双赢”都有正确的理解,则可能缩短这个进程。此时形成合作比在任务未开始前就形成动态联盟会容易得多,风险也小得多,因为合作的内容已经比较具体。这种竞标、协商即谈判可以用目前互联网支持的商品软件加上一些二次开发即可实现。第 6 章中将会介绍一个在众多竞争中选择合作伙伴的平台。

8) 由于服务请求方与服务提供方之间只有法律的约束,所以在选择提供方,即合作伙伴时对可信度的各个要素(包括历史信誉、当前的经营状态、专业技术水平、合作关系好坏等,这些都属于市场信息)的权重认真研究;对于合同的条款也必须有全面的考虑,要对各种可能发生的情况做出明确的规定。第 5 章的附件,介绍了一个这样的合同的样本。

## 第4章 互联网上合作设计的支撑技术

互联网上的合作设计能够提高企业对市场的快速响应能力,并且有利于各合作单位更加专业化,增强各自的实力,所以从根本上来说,互联网上的合作设计是企业实现快速、低成本产品设计的最佳途径。互联网上合作设计的发展,取决于多方面支撑技术的发展。由于涉及的技术多种多样,而且日新月异,这里只能选择其主要支撑技术,分为群体合作技术,产品设计信息共享技术,设计知识资源的发布、发现和集成技术和设计过程管理技术四个方面加以介绍。

### 1 群体合作技术

#### 1.1 CSCW 研究的发展

计算机技术和通信技术的发展,特别是近年来互联网的蓬勃发展和普及,为网上合作设计提供了技术保证,使网上合作设计成为可能。网上合作设计是群体合作活动的重要表现,它与 CSCW (Computer Supported Cooperative Work, 计算机支持的合作工作)的研究有着密切的联系。

CSCW 的研究最早始于 20 世纪 60 年代,并诞生了第一个实验系统 NLS/AUGMENT。20 世纪 70 年代中期,美国 Stanford AILab 建立了一个支持视频、声音、文本、图像等多种媒体的 CSCW 环境。到了 20 世纪 80 年代,和 CSCW 相关的计算机技术、网络技术、多媒体技术、数据压缩与存取技术、通信技术、分布与并行处理技术等都有了较大的发展,而且指导多媒体技术和 CSCW 技术的人机交互 (Human Computer Interaction, HCI) 理论的成熟,促进了 CSCW 的发展<sup>[1]</sup>。1984 年,来自麻省理工学院 (Massachusetts Institute of Technology, MIT) 的 Iren Grief 和数据设备公司 (Data Equipment Company, DEC) 的 Paul Cashman 两人组织了由来自不同领域的研究者组成的一个研讨会,共同讨论技术如何来支持人们协同工作的问题<sup>[2]</sup>。计算机支持的合作工作 (CSCW) 的术语就是在该研讨会中提出的。CSCW 的研究涉及到计算机科学、心理学、人类工程学、认知科学、社会学等多个学科。其目的在于建立一个基于计算机的协同工作环境,在此环境中人们可以相互合作,共同协调来完成同一任务。

经过大约 20 年的发展,CSCW 已经改变并且仍在

改变人们在个人以及组织之间合作和共享信息的方式。CSCW 最初来自两个研究领域:人机交互 (HCI) 和信息系统 (Information Systems, IS),前者主要关注通信和交互问题,后者主要关注协调问题和工作流技术。现在 CSCW 的许多概念和方法已经影响了许多研究领域的发展。对于产品设计理论和方法而言,CSCW 加快了产品开发和决策过程,在并行工程、合作设计、集成化产品开发和全面质量管理方面都有应用。

#### 1.2 CSCW 研究的内涵

Huber<sup>[3]</sup>认为 CSCW 是在与决策相关会议中支持群组决策的软件、硬件和语言组件和程序的集合。Brinck<sup>[4]</sup>认为 CSCW 是关于人们如何使用计算机技术来一起工作的研究,典型的主题包括电子邮件,包含其他用户认知活动的超文本和实时共享程序 (如合作写作或者绘画) 的使用。Willson 提出 CSCW 是一般性专有名词,它包含了对人们如何在群组中工作的方式的理解以及计算机网络和相关的硬件、软件、服务的驱动技术。Leslie<sup>[5]</sup>总结并提出 CSCW 系统要确保成功,应当包括以下特征:

- 1) 群组成员之间在同步和异步模式中的交互;
- 2) 群组成员执行的不同任务之间的协调;
- 3) 使人们能够从远距进行交互的装置;
- 4) 群组成员对数据的可视化以及可达性;
- 5) 群组成员之间数据、工程图样以及程序等的共享。

#### 1.3 CSCW 和群件的关系

许多研究者从不同的方面来定义群件,其中一些定义如下:

- 1) 群件是提高群组生产率的任何技术。(Lloyd, 1994)<sup>[6]</sup>
- 2) 群件本质上是一类软件,其能够从不同位置跨越计算机平台来共享系统应用和文件。(Benjamin, 1995)<sup>[7]</sup>
- 3) 群件是软件系统,通过提供相同共享环境和信息来支持从事共同决策任务的一组决策者。(Bidgoli, 1996)<sup>[8]</sup>
- 4) 群件是使能技术,它关注广泛的领域,包括合

作、人机交互和通过数字媒体的人人交互,进而为组织带来实质的改进和组成。(Korzeniowski, 1997)<sup>[1]</sup>

5) 群件,或协作软件,适合定义为应用程序,其允许在网络环境下共享、管理和控制信息。(Byrne, 1997)<sup>[6]</sup>

6) 群件是促进生产力的软件,它使人们能够交互,并且允许跨越时间和位置的合作。(Anthony, 1998)<sup>[11]</sup>

从群件定义的变化中,能够发现群件的内涵是从支持合作的广泛科技到基本上的纯软件技术。所以,群件基本上是一系列计算机软件系统,其功能是为人们的合作提供手段。

许多应用程序和系统,例如 20 世纪 70 年代出现的电子信息交换系统(Electronic Information Exchange System, EIES)和自动指导操作程序设计逻辑(Programmed Logic for Automated Teaching Operations, PLATO),20 世纪 80 年代出现的集成套件和办公自动化软件,如 Lotus Notes,20 世纪 90 年代出现的电子邮件, Mosaic 浏览器, Novell Groupwise, Microsoft Exchange, Lotus Domino 等推动了群件从初期到目前相对成熟的阶段。

群件经常被认为与 CSCW 是相同的含义,但事实并非如此。实际上群件(Groupware)比 CSCW 这个专有名词更早出现。Brinck 认为群件经常用于指示人们支持共同工作的技术,而 CSCW 是指研究这些技术的应用的领域。

所以, CSCW 位于理论层次,它关注的是人们如何工作和群件如何影响群体行为的研究和理论;群件位于技术和应用层次,它由支持个人和组织之间的群体合作的计算机软件系统组成。

## 2 产品设计信息共享技术

### 2.1 STEP 技术

国际标准化组织制定的 ISO 10303,即产品模型数据交换标准(Standard for Exchange of Product Model Data, STEP)是一套描述产品整个生命周期中产品数据的标准。它提供了描述产品数据的中性机制,能够保证在产品全生命周期中做到信息共享或交换而不丢失主要信息<sup>[12]</sup>。STEP 不仅能够描述几何信息,而且还包括参数化数据、特征、非几何数据,比如公差和表面粗糙度和加工计划等。STEP 还将进一步扩展以包含设计规划和加工知识等高层次的设计信息。因此,STEP 有希望成为工业界普遍采用的产品交换和数据共享的标准。

产品信息的交换,应该包括信息的获取、传输、存

储等。STEP 提供的数据交换方式有文件交换、应用程序界面、数据库实现和知识库实现。文件交换方法利用 ASCII 码或二进制文件,提供对产品数据描述的读写操作。STEP 自己定义了标准的中性文件格式,通过中性文件来实现产品信息交换。应用程序界面方法允许应用程序通过标准数据存取接口(Standard Data Access Interface, SDAI)来存取产品数据。SDAI 提供了一组完整的函数和子程序,以帮助程序员完成对数据的操作。数据库实现方法通过标准交换格式,调用标准存储软件 and 标准数据操作语言,如结构化查询语言(Structured Query Language, SQL)等,来实现对数据库中数据的读、写和修改。知识库实现方法通过知识及规则推理机制,实现与知识库管理系统的变换。

STEP 提供的建模语言是 EXPRESS,它用于描述产品全生命周期中数据所涉及的对象、对象所具有的信息单元以及对对象的限制与许可操作。EXPRESS 具有面向对象技术的继承机制,还有丰富的数据类型,因此具有很强的信息表达能力。

EXPRESS 的主要特点有:

- 1) 语言不仅能够为人所理解,而且能被计算机处理,描述的形式化使计算机自动检查和处理的可能性得到提高。
- 2) 语言能够区分 STEP 涉及的复杂内容。
- 3) 语言的重点放在实体(Entity)的定义上,实体的定义包括实体的属性和施加的约束条件。
- 4) 语言与具体的实现无关。
- 5) EXPRESS 语言已经历了国际化的进程,成为能够满足工业需求的标准语言。

由于 EXPRESS 不是编程语言,因此需要将 EXPRESS 描述的信息模型映射到具体的编程语言,才能由计算机应用系统进行数据处理,即使用时要解决 EXPRESS 与编程语言(C++, Java 等)之间的联编。

与 EXPRESS 相关的另外两个概念是 EXPRESS-G 和 EXPRESS-X,EXPRESS-G 是 EXPRESS 的一个子集,是 EXPRESS 的图形化描述。它采用符号或不同形式的标志图符以表示模型中的主要条目对象。作为产品全生命周期中数据的表达和交换标准,STEP 具有完整性和确定性。因此,STEP 标准非常庞大,包含的细节信息非常多。在实际应用系统中,许多细节信息并不一定需要,因此需要对产品模型进行“简化”。而且,经过“简化”的产品模型,更容易被人理解并实现。但是,不同应用系统对同一产品模型的视角不同,要求“简化”的方式也不同。因此,需要提供一种途径,能很容易地获得不同应用系统需要的“简化”的产品模型,即“视”模型(View Information Model)。

EXPRESS-X 语言作为对 EXPRESS 语言的扩充,包含了一些结构,可以方便地获得对“视”模型的定义。EXPRESS-X 语言用来描述不同的 EXPRESS 模式中的实体之间的映射,这种映射可用来创建一个数据集的不同表达。EXPRESS-X 语言的任务就是描述用 EXPRESS 语言建立的信息模型之间的映射。EXPRESS-X 目前还没有成为国际标准,处在不断的发展和完善中。

## 2.2 XML 技术

可扩展标记语言 (Extensible Markup Language, XML) 是一种可扩展的和自描述的标记语言。它由万维网协会 (World Wide Web Consortium, W3C) 创建,用来克服超文本标记语言 (Hypertext Markup Language, HTML) 的局限。和 HTML 一样,XML 基于标准通用标记语言 (Standard Generalized Markup Language, SGML)。尽管 SGML 在出版业已使用了数十年,但其过于繁杂,限制了其广泛使用。

XML 并不是 HTML 的替代品,XML 和 HTML 有不同的用途。它们的主要区别在于 XML 是用来描述数据,并集中解决数据是什么的问题,而 HTML 是用来显示数据,并集中解决数据如何显示的问题。

XML 是一套定义语义标记的规则,这些标记将文档分成许多部件并对这些部件加以标识。用户可以定义自己需要的标记。这些标记必须根据某些通用的原理来创建,但是在标记的意义上,也具有相当的灵活性。XML 定义了一套元句法,与特定领域有关的标记语言 (如 MusicML、MathML 和 CML) 都必须遵守。如果一个应用程序可以理解这一元句法,那么它也就自动地能够理解所有的由此元语言建立起来的语言。浏览器不必事先了解多种不同的标记语言使用的每个标记。XML 标记描述的是文档的结构和意义。它不描述页面元素的格式化。可用样式单为文档增加格式化信息。文档本身只说明文档包括什么标记,而不是说明文档显示是什么样的。

XML 的主要特点是<sup>[12]</sup>:

1) 自描述性:XML 文档是规格清晰的文档,界定 XML 内容的标记会给所界定的数据中的每一个元素命名,在标记中还可以通过特定的属性为所描述的元素提供某些附加信息,这样可以清楚理解 XML 文档的内容,而且计算机也能够自动地理解数据的含义;

2) 可扩展性:XML 的扩展性体现在其能够定义新的标记及作为元标记语言能够定义其他用途的标记语言或者标准;

3) 可校验性:用户可以通过文档类型定义 (Document Type Definition, DTD) 或 XML 架构

(XML Schema) 来校验 XML 文档的格式是否满足 DTD 或 XML 架构的约束;

4) 层次结构:能够保持信息的层次性描述;

5) 丰富的链接定义:对应于 HTML 的单向单通道链接,XML 提供各种不同链接,如一对多、多对一和双向链接;

6) 多样的样式表支持:XML 把数据内容与数据的表现形式分开。这样既可以只关心数据的逻辑结构,也可以通过样式表来格式化数据的表现。

XML 文档可以表示结构化数据,可作为文本数据库存储数据,也可作为行业中数据交换的标准表示。这些都需要对 XML 文件的数据进行描述,如数据类型、长度等,在 1998 年发布的 XML 1.0 规范第一版中,就是采用 DTD 完成数据建模工作的。但是,由于 DTD 是基于正则表达式的,描述能力有限;没有数据类型的支持,在大多数应用环境下能力不足;约束定义能力不足,无法对 XML 文档做出更细致的语义限制;DTD 结构不够结构化,重用的代价相对较高;DTD 不是使用 XML 作为描述手段,其构建和访问没有标准的编程接口。所以 W3C 在 2001 年推出 XML Schema 标准,它目前已经基本取代了 DTD 在 XML 环境下数据建模的地位。XML Schema 是针对 DTD 的缺点而设计的,它是完全使用 XML 作为描述手段,具有很强的描述、扩展和处理维护能力。

XML 是描述数据和交换数据的普遍接受的标准化的方法,同时它也支持互联网上的合作设计的支撑技术、工具和应用提供了描述数据和交换数据的标准、普遍应用的方法。

## 3 设计知识资源的构建、发布、发现和集成技术

### 3.1 TCP/IP 协议系列

TCP/IP 是传输控制协议 (Transmission Control Protocol, TCP) 和网际协议 (Internet Protocol, IP) 的简称,它是 Internet 采用的协议标准,也是全世界采用的最广泛的工业标准。事实上,它是一个协议系列,目前包含了 100 多个协议,用来将各种计算机和数据通信设备组成实际的计算机网络。这个协议系列的正确名字应是 Internet 协议系列,而 TCP 和 IP 是其中的两个最基本和最重要的协议,也是广为人知的,因此通常用 TCP/IP 来代表整个 Internet 协议系列。其中有些协议是为很多应用需要而提供的低层功能,包括 IP、地址解析协议 (Address Resolution Protocol, ARP)、逆向地址解析协议 (Reverse Address Resolution Protocol, RARP)、网间控制报文协议

(Internet Control Messages Protocol, ICMP)、网间组管理协议 (Internet Group Management Protocol, IGMP)、TCP 以及用户数据包协议 (User Datagram Protocol, UDP) 等; 另一些协议则在底层基础上完成特定的应用任务, 如 Telnet 协议实现远程登录, 文件传输协议 (File Transfer Protocol, FTP) 完成文件传递, 简单邮件传输协议 (Simple Message Transfer Protocol, SMTP) 完成电子邮件发送, 域名系统 (Domain Name System, DNS) 实现域名解析服务等。

TCP/IP 协议系列遵守一个四层的模型概念: 应用层、传输层、互联层和网络接口层, 各层包含的网络协议如表 49.4-1 所示。

表 49.4-1 Internet 协议的分层

应用层	Telnet, FTP, SMTP, HTTP, DNS 等
传输层	TCP, UDP
网络层	IP, ARP, RARP, ICMP, IGMP
网络接口层	Network Interface (Physics Networks)

在 Internet 协议栈中, 超文本传输协议 (Hypertext Transfer Protocol, HTTP) 位于 TCP/IP 协议栈上层。HTTP 协议是互联网上进行数据传输的基本协议, 它是 Internet 协议栈中重要的协议。Web 服务器和浏览器均采用 HTTP 协议传输由超媒体组成的 Web 文档。另外 HTTP 还能维持多媒体信息的完整性。可以说 HTTP 是互联网上的图像、音频、视频、超文本等信息的传输载体。互联网之所以把 HTTP 当作其基本协议是因为没有其他的协议能提供如此全面的性能。HTTP 是一个基于消息的协议。在 HTTP 中有两部分消息, 一部分是从浏览器 (客户端) 发往服务器的请求, 另一部分是服务器对客户端的响应。

HTTP 分四步完成一次事务:

- 1) Web 浏览器和服务器之间建立 TCP/IP 连接;
- 2) 浏览器 (客户端) 向服务器发出请求;
- 3) 服务器响应客户端的请求;
- 4) 客户端与服务器断开连接。

HTTP 本身也在不断发展, 在其基础上产生了安全超文本传输协议 (Secure Hypertext Transfer Protocol, S-HTTP) 协议。S-HTTP 协议也是一个基于可靠传输协议 TCP 的高层网络协议, 它通过提供一系列 HTTP 客户和服务器之间的安全机制, 从而对 Web 事务提供独立的安全服务, 如保密、鉴别/完整性保护以及源点的无否认等来保护 Web 浏览器/服务器在 Internet 上传送敏感信息, 同时保持了 HTTP 的交互模式和实现特点。

### 3.2 分布式对象技术

传统的面向对象的技术通过封装、继承及多态提

供了良好的代码重用功能, 但是这些对象只存在于程序内部中, 外面的世界并不知道它们的存在, 也无法访问它们。为了更好地实现应用共享, 人们提出了分布式对象的概念。

根据面向对象的知识, 一个对象可以封装变量和函数。在常规的环境下, 程序在某个进程中创建对象的实例, 访问成员变量和调用成员函数。如果对某个对象的调用和此对象的实现位于不同的计算机的不同进程中, 这个对象称之为分布式对象。

分布式对象存在于网络的任何地方, 可被远程客户以方法调用的形式访问。客户无需知道它所访问的分布式对象在网络中的具体位置和使用的操作系统。这就需要提供 一个标准的构件框架, 使不同厂家的软件通过不同的地址空间、网络和操作系统交互访问。目前分布式对象技术的三大主流为对象管理组织 (Object Management Group, OMG) 的 CORBA/IIOP (Common Object Request Broker Architecture/Internet Inter-ORB Protocol, 公共对象请求代理体系结构/因特网对象请求代理间协议)、Microsoft 的 ActiveX/DCOM (Distributed Component Object Model, 分布式组件对象模型) 和 Sun 的 Java/RMI (Remote Method Invocation, 远程方法调用)。

OMG 是 CORBA 规范的制定者, 由 800 多个信息系统供应商、软件开发者和用户共同构成的国际组织, 建立于 1989 年。OMG 在理论上和实践上促进了面向对象软件的发展, 它只制定规范, 而不提供具体的实现。通过 CORBA 规范的制定, OMG 将对象引入分布式的环境中。CORBA 是 OMG 进行标准化分布式对象的计算基础。CORBA 自动匹配许多公共网络任务, 例如对象登记、定位、激活、多路请求、组帧和错误控制、参数编排和反编排、操作分配等。图 49.4-1 是 OMG 对象服务参考模型结构<sup>[4]</sup>。

其中, 公共对象服务包括了支持分布式系统正常工作的各类基本的系统级服务; 公共设施包括支持分布式系统高效开发和有效工作的各类面向领域的常规服务和工具; 应用对象涉及各种应用软件, 它在公共对象服务和公共设施的支持下完成相应的应用逻辑; ORB 就像是一条软件总线, 它把分布式系统中的各类应用和对象连接成相互作用的整体。

组件对象模型 (Component Object Model, COM) 是微软制定的一套软件开发技术, 它定义了一组应用编程接口 (Application Programming Interface, API) 和二进制标准, 让来自不同编程语言和平台的彼此独立的对象互相通信。Microsoft 推出 ActiveX 的目的是为了使开发人员能够把计算机桌面环境与构成 Internet 及其大量资源的环境集成起来, 同时保护在

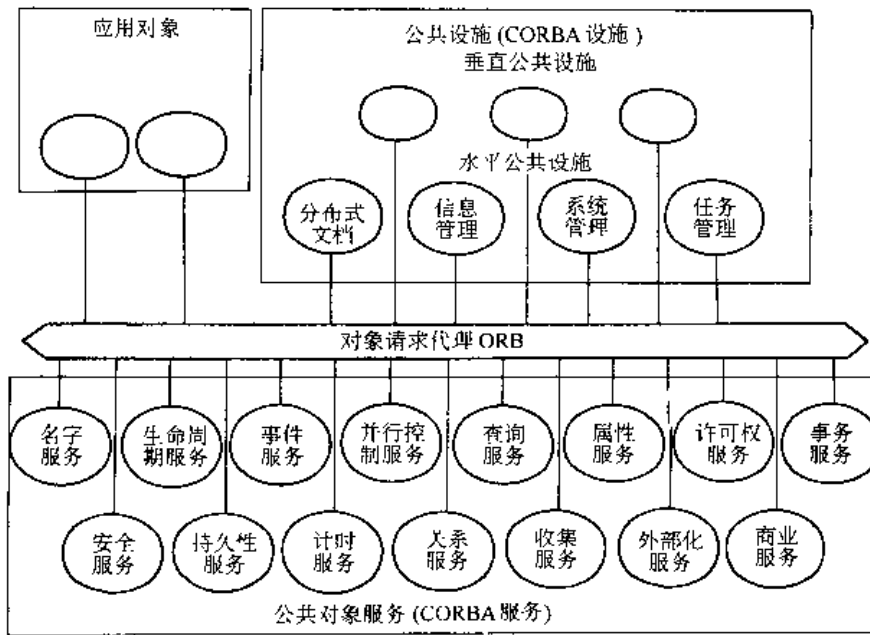


图 49.4-1 OMG 对象服务参考模型结构

Windows 中现有的开发投资。ActiveX 包括对两个现有 Microsoft 技术：Win32API 和组件对象模型 COM 的一系列扩充和增强。ActiveX 简化了对象链接和嵌入 (Object Linking and Embedding, OLE) 的标准接口；并且通过永久链接、永久嵌入等修改了 OLE 对数据和特性的管理，从而使得 ActiveX 小且富有效率。通过 ActiveX 技术，开发人员就能够把可复用的软件模块组装到应用程序或者服务程序中去了。DCOM 扩展了 COM 技术，使其能够支持在网络上不同计算机的对象之间的通信。

Java 是一个应用程序开发平台，它提供了可移植、可解释、高性能和面向对象的编程语言及运行环境。RMI 是分布在网络中的各类 Java 对象之间进行方法调用的 ORB 机制。因为 Sun 公司是 OMG 的创始成员，CORBA 中的许多内容是以 Sun 提交的方案为核心制定的，因此 CORBA 与 Java 有着天然的联系。CORBA 与 Java 的主要区别在于<sup>[15]</sup>：

1) CORBA 遵循与程序设计语言的无关性，而 Java/RMI 依赖于 Java 和 Java 虚拟机；

2) Java/RMI 能够使对象在 Internet 上迁移和执行，而 CORBA 2.0 中只考虑对象的远程访问，没有对象作为“值”传递的承诺。

CORBA 和 ActiveX/DCOM 的主要区别在于：

1) CORBA 依赖于 IIOP 进行远程对象通信，DCOM 则依赖于对象远程处理过程调用以达到相同的目的；

2) CORBA 体系结构是基于对象请求代理的；DCOM 则以 COM 作为它的基础，事务处理则依赖于

微软事务处理服务器 (Microsoft Transaction Server, MTS) 或微软消息队列 (Microsoft Message Queue, MSMQ)；

3) CORBA 规范不是针对特定厂商的，因此 CORBA 应用能运行于不同的硬件平台上。ActiveX/DCOM 则是由微软制定和拥有的体系结构，并且基本上只能运行于微软操作系统支持的硬件平台上；

4) CORBA 提供跨平台支持，ActiveX/DCOM 则局限于微软操作系统，当然，现在 Macintosh 也开始提供对 COM 的支持，某些 Unix 也在逐步提供对该技术的支持；

5) 基于 CORBA 开发出来的中间件对大多数企业而言过于复杂，不易使用。而且 CORBA 规范由许多厂商一起创立，而每个厂商都有自己的定义对象的方式和方法，这便使互通性产生了一定问题；而对于 ActiveX/DCOM 在易用性方面相对要比较好。

### 3.3 Web Services 技术

虽然采用 CORBA, ActiveX/DCOM 与 Java/RMI 等技术的系统具备高度的可靠性与安全性，但是它们在几个关键性问题上，例如消息传送通信协议，服务功能的抽象描述，服务在网络上的发布与发现机制等均采用了封闭的、私有的解决方法，这导致开发这些系统的成本极为高昂，而且造成了分布式应用之间新的割裂。

以可扩展标记语言 (XML)、简单对象访问协议 (Simple Object Access Protocol, SOAP)、统一描述、发现和集成协议 (Universal Description, Discovery and Integration, UDDI) 和 Web 服务描述语言 (Web

Service Definition Language, WSDL) 为核心标准的 Web 服务的出现, 为解决分布的知识资源应用的构建、发布、发现和集成问题提出了新的解决途径。

SOAP 是一种简单的、轻量级的基于 XML 的机制, 用于在网络应用程序之间进行结构化数据交换。SOAP 包括三部分: 一个定义描述消息内容的框架的信封、一组表示应用程序定义的数据类型实例的编码规则, 以及表示远程过程调用 (Remote Procedure Calls, RPC) 和响应的约定。SOAP 可以和各种网络协议 (如 HTTP、SMTP、FTP、MQ 或者 IIOP 上的 RMI) 相结合使用, 或者用这些协议重新封装后使用。

Web 服务是描述一些操作 (利用标准化的 XML 消息传递机制可以通过网络访问这些操作) 的接口。Web 服务是用标准的、规范的 XML 概念描述的, 称为 Web 服务的描述。这一描述囊括了与服务交互需要的全部细节, 包括消息格式 (详细描述操作)、传输协议和位置。该接口隐藏了实现服务的细节, 允许独立于实现服务基于的硬件或软件平台和编写服务所用的编程语言使用服务。这允许并支持基于 Web 服务的应用程序成为松散耦合、面向组件和跨技术实现。Web 服务履行一项特定的任务或一组任务。Web 服务可以单独或同其他 Web 服务一起用于实现复杂的聚集或商业交易。

Web 服务的用户 (或者用户程序) 可以采用 UDDI 协议, 来发现 Web 服务供应商发布的 Web 服务; Web 服务采用 WSDL 语言确定服务的接口定义, 抽象描述服务的功能; 用基于 SOAP 协议的 XML 文档通过 HTTP、FTP 和 SMTP 等常用通信方式交换数据, 传递消息等。Web 服务体系架构模型见图 49.4-2。

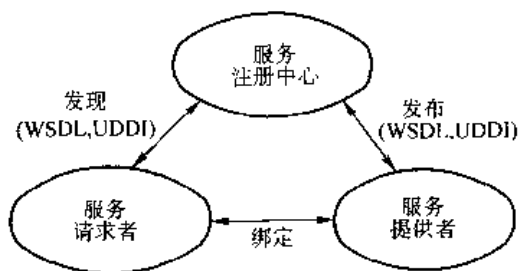


图 49.4-2 Web 体系架构模型  
Web 服务体系架构模型包括:

1) 服务提供者。从使用者的角度看, 这是服务的拥有者。从体系结构的角度看, 这是托管访问服务的平台。

2) 服务请求者。从提供者的角度看, 这是要求满足特定功能服务的使用者。从体系结构的角度看, 这是寻找并调用服务, 或启动与服务交互的应用程序。服务请求者角色可以由浏览器来担当, 由人或无用户界面的程序 (例如另外一个 Web 服务) 来控制它。

3) 服务注册中心。这是可搜索的服务描述注册中心, 服务提供者在此发布他们的服务描述。在静态绑定开发或动态绑定执行期间, 服务请求者查找服务并获得服务的绑定信息 (在服务描述中)。对于静态绑定的服务请求者, 服务注册中心是体系结构中的可选角色, 因为服务提供者可以把描述直接发送给服务请求者。

Web 服务的优点在于:

1) 互操作性。由于 SOAP 协议的存在, 任何 Web 服务都可以与其他 Web 服务进行交互。这样就避免了在 CORBA, ActiveX/DCOM 和其他协议之间转换的麻烦。

2) 普遍性。Web 服务支持使用 HTTP 协议和 XML 进行通信。因此, 任何支持这些技术的应用和设备都可以拥有和访问 Web 服务。

3) 易用性。Web 服务的概念易于理解, 任何开发语言都可以用来编写 Web 服务。

4) 开放性。ActiveX/DCOM 和 CORBA 等是以不同公司和组织支持的协议和标准为基础的, 而 Web 服务是建构在四个开放的核心标准和协议之上, 得到了普遍的支持。

Web 服务也可以认为是封装成单个实体并发布到网络上以供其他程序使用的功能集合, 是用于创建开放分布式系统应用的构件。Web 服务是一类软件, 独立于硬件, 同时与操作系统、网络和数据库等软件系统也不同; 实现分布式应用之间的互连和互操作; 解决了网络通信的问题。所以 Web 服务技术也是分布式对象理论和应用的发展。

### 3.4 UDDI 技术

统一描述、发现和集成协议 (UDDI) 是一套基于 Web 的、分布式的、为 Web 服务提供的信息注册中心的实现标准规范, 同时也包含一组使企业能将自身提供的 Web 服务注册以使别的企业能够发现的访问协议的实现标准。

UDDI 注册使用的核心信息模型由 XML Schema 定义。使用 XML 是因为它提供了平台无关的数据描述并很自然地描述了数据的层次关系。而选择 XML Schema 是因为它支持丰富的数据类型, 方便描述方式及其按信息模型对数据进行验证的能力。

UDDI XML Schema 定义了六种主要信息类型, 如图 49.4-3 所示, 它们是技术人员在需要使用合作伙伴所提供的 Web 服务时必须了解的技术信息。它们是: 商业实体信息 (businessEntity)、服务信息 (businessService)、绑定信息 (binding Template)、服务调用规范 (tModel)、关联关系断言 (publisher Assertion) 的说明信息和实体订阅信息 (subscription)。



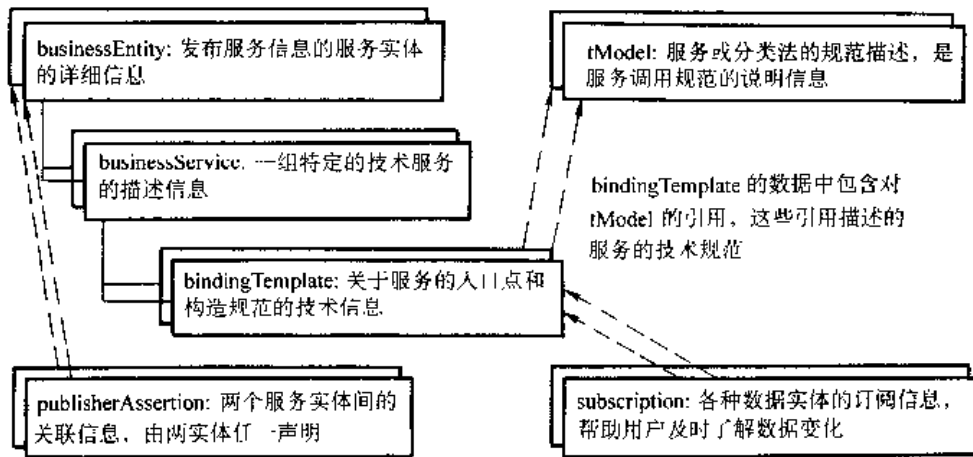


图 49.4-3 UDDI 信息模型结构图

UDDI 的提出最初是为了解决在电子商务中购买者、供应商、交易市场和提供者无法互相了解各自的需求与能力, 以及无法在 Internet 上进行应用和服务的集成的问题。在产品领域, 尤其是合作产品设计领域, 通过基于设计实体和资源单元的服务请求和服务提供的知识服务来合作完成产品设计任务, 同样存在如何在 Internet 上了解各自的需求和能力的问题, 同样也需要在不同的应用和服务中进行集成。

分析知识服务发布和发现问题, 可以得到从传统模式逐渐发展到互联网模式的 5 种模式, 即:

1) 如果在“信息都是已知”的基础上, 也就不存在知识发现等的困难。这也是合作总是最先寻找最熟悉的合作伙伴的原因。但是对于互联网上的合作设计, 就需要逐步过渡到寻求一种机制来发现目前不熟悉, 但是能够完成知识服务的最优的智力资源单元。

2) 传统上发布机制, 是编写电话黄页, 按行业、职业等标准分类收集、整理电话号码和详细地址; 而传统上的发现机制, 是查询黄页, 使用电话和每个潜在的合作伙伴进行联系, 然后找出合适的对象。而对于一个提供知识服务的知识服务实体来说, 需要配备具备相当

技术能力的专业人员去满足这样随机的服务发现需求显然是不合适的。

3) 目前互联网上的知识服务的发布也可以通过自己把服务的内容等信息注册到某些通用的互联网搜索站点 (如 Google, Altavista, Lycos 等), 或者由搜索网站来通过某些网络爬虫程序来搜索网页, 分类整理; 知识服务的发现是通过使用者使用搜索网站上的搜索引擎进行搜索。一方面由于互联网上的各个服务器中的网页基本上都是高度无组织化的信息, 另一方面能够提供知识服务的智力资源单元信息与普通的文档信息有较大的区别, 通过搜索引擎的优化来高效的完成知识服务发布和发现的工作是比较困难的。

4) 也有研究者提出在知识服务供应商的每个网站上放置一个知识服务描述文件。这样, 至少那些网络爬虫程序能够发现并为它们建立索引。可是这种定位知识服务的方法完全依赖爬虫程序的能力。这种分布式的机制是可扩展的。但它缺少一种机制来保证服务描述格式的一致性, 也不能便捷的跟踪不断发生的变化。

5) 基于 UDDI 协议来实现知识服务中的知识发布、知识发现的具体过程如图 49.4-4 所示。

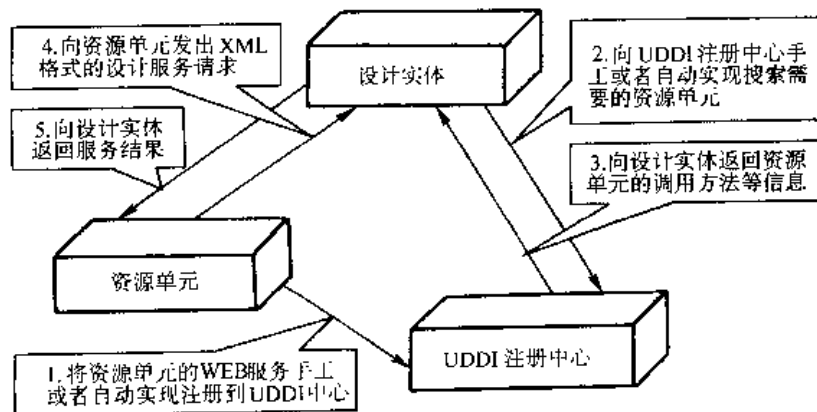


图 49.4-4 基于 UDDI 协议的知识发布与知识发现的过程



a. 资源单元在自己的 Web 服务器上提供产品设计的 Web 服务, 并手工或者编程实现自动在公有的 UDDI 注册中心或者专有的产品设计 UDDI 注册中心进行注册。

b. 设计实体手工或者编程实现自动向 UDDI 注册中心提出搜索请求, UDDI 注册中心在全球范围搜索需要的资源单元。

c. UDDI 注册中心将搜索到的资源单元的调用方法等信息返回给设计实体。

d. 设计实体按照返回的资源单元的调用方法等信息, 向资源单元的 Web 服务发出 XML 格式的服务请求。

e. 资源单元完成请求的任务后, 向设计实体以 XML 格式返回需要的服务结果。

以上过程可以由计算机自动完成, 当然人在中间也起重要作用, 可以手工干预和识别搜索过程。在实际设计过程中, 设计实体(服务请求者)和资源单元(服务提供者)也可以存在于一体, 但这并不妨碍各自具有自己的特征和在自己特征支配下行为。

### 3.5 智能体 Agent

#### 3.5.1 Agent 的基本概念

自从 Agent 技术从 20 世纪 70 年代出现在人工智能领域, 几十年来, Agent 技术不仅在人工智能领域, 特别是分布式人工智能领域 (Distributed Artificial Intelligence, DAI) 有了长足的发展, 而且向软件工程领域、人机交互系统领域、并行工程 (Concurrent Engineering, CE) 领域、计算机支持的协同工作 (CSCW) 领域、工作流管理系统 (Work Flow Management System, WFMS) 领域、决策支持系统 (Decision Support, System, DSS) 领域、CAX (CAD/CAE/CAM/CAPP) 领域等做多方面的渗透, 在促进这些领域理论和实践的发展的同时, 也有力的促进了自身的发展。近年来, Agent 的研究和开发得到了迅速的发展。它所具有的自身的社会性、开放性、自主性和智能性已经使其成为计算机科学的一个非常重要而活跃的领域。同时, 基于 Agent 的系统已经用于工业、商业、游戏、医疗保健、军事等领域的应用系统开发。Agent 技术可以认为是分布对象技术的一种增强和扩展, 它将为网上合作设计提供更加有利的支持。

Agent 的出现和发展为相关领域带来了新的方法, 但是关于 Agent 目前还没有一种能为大家所共同接受和认可的定义。在不同的领域, 对于不同的对象和场合存在极大的差异。Agent 在分布式人工智能领域, 侧重于主体概念, 可以认为是在还有其他 Agent 存在

的情况下, 能够连续、自主地处理所处环境中发生的事件的功能的总合。在这里所说的“自主”是指系统中 Agent 工作时不要求由人经常不断地引导和干预; 从认知角度来看, Agent 是一个由信念、能力、选择、承诺等功能部件组成的实体; 在计算机科学中, Agent 像一个自我包含的并行执行软件过程一样, 它封装某种状态, 通过消息的传送与其他的 Agent 进行通信; 在软件工程, 软件开发模型中, 基于 Agent 计算 (Agent-Based Computing, ABC) 被认为是软件开发的下一个重要突破, 面向 Agent 的程序设计 (Agent-Oriented Programming, AOP) 也被认为是面向对象程序设计 (Object-Oriented Programming, OOP) 的一次革命; 在实际应用领域中, Agent 又往往是与它的物理表征密切相关的<sup>[16]</sup>。Agent 在工厂常常被看成是“机器 Agent”, 在大型计算机网络中则被看作是软件机器人 (Softbots), 为完成某种特定任务的机器人还被称为任务机器人 (Taskbots), 另外还有知识机器人 (Knowbots)、用户机器人 (Userbots)。Agent 在系统中可以不同的身份起着不同的作用, 比如搜索 Agent、报告 Agent、导航 Agent、管理 Agent、开发 Agent、分析和设计 Agent、测试 Agent、包装 Agent 和辅助 Agent 等。同时, Agent 与其他的软件或硬件不同, 它往往出现在系统的“高层”。所谓的“高层”, 明显表现在 Agent 使用符号表示和它具有一些类似认识的功能, 比如, Agent 具有很强的通信能力, 除了刺激-反应规则之外, Agent 还可以包含有符号规划, 甚至 Agent 还具有使用自然语言的能力。这也就是说基于 Agent 的应用操作是在知识层, 而不是像分布式计算那样在符号层。

由于各领域对于 Agent 目前还没有一种能为大家所共同接受和认可的定义, 这里给出一些没有与任何特定领域联系的更为一般的 Agent 概念定义。有人认为, Agent 是指驻留在某一环境下能持续、自主地发挥作用, 满足反应性、社会性、主动性等特征的计算实体<sup>[17]</sup>; 也有人认为, Agent 就是能够准确完成用户指定任务的软件和/或硬件设备; 还有人从目的、属性与主体关系三个方面给出一个 Agent 的定义: Agent 是一个运行于动态环境中的具有较高自制能力的实体 (即自制体, 可以是系统、机器等; 软件 Agent 是一个计算机软件程序), 其根本目标是接受另外一个实体 (即主体, 可以是用户、计算机程序、系统或机器等) 的委托并为之提供帮助和服务, 能够在该目标的驱动下主动采取包括社交、学习等手段在内的各种必要的行为, 以感知、适应并对动态环境的变化进行适当反应。它与其服务主体之间具有较为松散和相对独立的关系。

### 3.5.2 Agent 的属性

根据 Agent 的应用, Wooldrige 将 Agent 划分为弱 Agent (Weak Agent, WA) 和强 Agent (Stronger Agent, SA) 两类。WA 通常用于硬件和基于软件的计算机系统, 一般具有下面的属性<sup>[18]</sup>。

1) Agent 是自主的和智能的 (Autonomy and Intelligent)。它能够在没有人或者在其他 Agent 的直接干预下, 对复杂的刺激进行响应并伴随产生内部状态的控制和适应性的行为。

2) Agent 具有社会能力 (Social Ability)。Agent 彼此之间能够以某种 Agent 通信语言进行通信, 尽管这种通信模式可能十分复杂。

3) Agent 具有很好的反应能力 (Reactivity)。Agent 能够理解它所处的环境。比如, 物理世界、图形界面、其他的 Agent、互联网或者它们的联合, Agent 都能够及时地对系统中所发生的变化做出适当的反应。

4) Agent 的活动具有主动性 (Pro-Activeness)。它们对环境不是简单地做出反应, 而是积极主动地朝着总体目标行动。

5) 一般情况下, Agent 分布在网络之中, 因此它们的行为既具有局部效应又具有全局效应。

在某些研究领域, 尤其是人工智能领域, 科学家们往往更关注 Agent 的智力行为, 从而构成了 SA 的概念。SA 与 WA 不同, 它除了 WA 所具有的属性外, 还强调计算机系统中人的智力和精神上、情感上的因素, 比如知识、信念、意图、拒绝、承诺等。

### 3.5.3 Agent 的优点、局限性和面临的挑战

Agent 的优点在于其能够解决空间分布的信息资源和专家知识的有效利用问题, 允许多个现有的智能系统相互联系和相互合作, 实现分布、并行、合作问题求解, 能够适应和容许某些不确定事件的发生, 增强系统容错和故障恢复能力; 通过不同数量和不同能力的 Agent 组成针对不同问题的 MAS, 提高程序代码的可重用性并增强系统的可扩展性和灵活性; 通过 Agent 的模块化增强系统的可维护性, 能够在模块或局部范围内处理异常事件而不致将其扩散到整个系统, 从而降低系统对异常事件的敏感性。因此, Agent 技术具有广泛的应用前景, 主要涉及复杂和并发系统的建立与管理、流动访问与管理、信息搜集与处理、分布计算与协同工作、电子商务以及用户界面和中间件等。

Agent 的局限性在于基于 Agent 系统并不是十全十美的, 它也有其不足和局限性以及不适合解决的问

题域<sup>[17]</sup>。当然这些局限性在实践应用中是可以采取某些方法来避免和克服的。

1) 无全局控制。在基于 Agent 系统中, 每个 Agent 都是自主的计算实体, 系统中没有一个全局控制的 Agent 用于协调和管理整个系统的活动。因此, 整个系统缺乏全局性控制。

2) 无全局的观点。由于 Agent 是自主的计算实体, Agent 的行为完全是由 Agent 根据其内部的状态来决定的。同时, 对于系统中的任一 Agent 而言, 要获得关于系统的完整性、全局性的知识是不可能的, 这就意味着 Agent 做出的行为决策只能是站在其自身的立场上, 是部分的而不是全局的, 其行为决策至多是部分最优而不是全局最优的。

3) 不可预测和不确定性。由于基于 Agent 系统中的各个 Agent 都是自主的计算实体, Agent 之间通过灵活的交互来实现问题的求解。Agent 间的交互作用不是通过在设计阶段就已经确定的接口进行, 而是在运行时通过 Agent 之间的合作、协同、协商、竞争等方式来实现。这意味着基于 Agent 系统的整体性质和行为具有不可预测性和不确定性, 人们很难预测一个任务是否最终能够得到完成, 也不能事先确定其他 Agent 能否友好地与其合作或者就某些事宜达成一致。

4) 无序和混乱。Agent 的高度自治性意味着 Agent 可以自主地实施行为, 如果不能对基于 Agent 的系统实施有效的管理, 那么就有可能导致系统的无序和混乱。

基于 Agent 的系统的开发而面临着许多问题和挑战:

1) 对 Agent 的认识问题。目前还没有一个能为大家所共同认可和接受的 Agent 定义, 不同领域的人员对它有不同理解和认识, 这使得研究工作不能收敛, 同时缺乏坚实的概念基础和具体、明确的目标, 从而在一定程度上影响了这一研究领域的发展和应。

2) 缺乏开发环境和编程工具的支持。目前大多数基于 Agent 的系统是利用非 Agent 技术来实现的, 这意味着 Agent 技术还不成熟以及 Agent 技术还没有真正为广大计算机工作者所认可和接受; 尽管人们已经提出了许多面向 Agent 的程序设计语言如 Agent-0, CONGOLOG, Concurrent METATEM 等等, 但至今还没有一种实用、为大家所广泛接受和使用的面向 Agent 的程序设计语言。

3) 缺乏系统开发方法的支持。目前我们还没有一种系统的方法用于指导基于 Agent 的系统的开发, 许多基于 Agent 的系统完全是借助于软件开发人员所具有的零散 (而不是系统)、具体 (而不是抽象)、甚至

是与 Agent 不相关的经验知识建立起来的。尽管人们已经提出基于 Agent 的软件工程的思想,但它远不及面向对象方法那样有一系列比较系统的手段用于指导基于 Agent 的系统的需求分析、设计、实现和验证等工作。

## 4 设计过程管理技术

### 4.1 PDM 技术

产品数据管理(PDM)技术最早出现于 20 世纪 80 年代初期,目的是为了解决大量工程图纸、技术文档以及 CAD 文件的计算机化的管理问题,后来逐渐扩展到产品开发中的三个主要领域:设计图纸和电子文档的管理,材料报表(Bill Of Materials, BOM)的管理以及与工程文档的集成,工程变更请求/指令(Engineering Change Request/Order, ECR/ECO)的跟踪与管理<sup>[19]</sup>。

由于 PDM 技术与应用范围发展很快,人们对它还没有一个统一的认识,给出的定义也不完全相同。CIMData 为 PDM 下的定义是:“PDM 是一门管理所有与产品相关的信息和所有与产品相关的过程的技术”<sup>[20]</sup>。而 Gartner Group 给出的定义为:“PDM 是在企业范围内从策划到产品构筑一个并行化协作环境(Concurrent Art-to-product Environment, CAPE, 由供应、工程设计、制造、采购、市场与销售、客户等构成)的关键使能器。一个成熟的 PDM 系统可使所有参与创建、交流、维护设计意图的人们在整个信息生命周期中自由共享与产品相关的所有异构数据,包括图纸与数字化文档、CAD 文件和产品结构等”<sup>[19]</sup>。所以,PDM 从狭义上讲,它仅管理与工程设计相关的领域内的信息;而从广义上讲,它可以覆盖到从市场需求、研究开发、产品设计、工程制造、销售、服务与维护等全生命周期中的信息。

产品数据管理(PDM)系统是进行设计过程管理的有效工具。PDM 系统确保跟踪设计、制造所需的大量数据和信息,并由此支持和维护产品。可见 PDM 的主要应用领域包括:数据存储和管理、工作流程管理、设计变更控制管理、产品结构与配置管理、标准件管理、图像管理和项目管理等,它是实现互联网上的合作设计的设计过程管理的有效工具。由于一个设计任务总是由若干子任务构成,而设计过程也总是由若干子过程构成,并且都具有层次结构,所以在分布式资源环境下的合作设计中,一个 PDM 的管理权限和范围应当与它的任务相一致。下级子任务的 PDM 不一定能管理上级子任务的所有事务,反之,上级子任务的 PDM 也不一定管理下级子任务的所有事务。下面所

讨论的合理化检查、设计过程记录、冲突检验以及更改通知等,都有这种情况。

为了保证设计出来的产品不仅可制造、装配,而且在性能上也完全符合设计要求,在互联网上的合作设计环境中应该包括设计合理化检查工具。设计合理化检查工具内含一个与此产品设计相关的智能系统。这个智能系统中存放有设计过程中应遵循的一般性原则和常识,以及本次产品设计的约束条件。在设计过程中,设计合理化检查工具将监视每一步的设计结果,并由智能系统对结果进行检查、判断。一旦发现有违背设计原则、设计约束和常识的情况出现,智能系统便立即告知设计师什么地方设计有不妥之处、其原因以及可能的解决方案。

在设计过程中,如果发现目前的设计不能满足设计要求,就有可能要采用或参考前面的设计过程中出现的某个方案,因此每个设计参与方必须拥有设计过程的记录工具。它将完整地记录下产品从开始设计到设计完成的每一步设计操作,以便设计人员对以往的设计结果进行查询、调用。

在合作设计的模型下,每个参与合作设计的单位负责产品某一部分的结构和功能设计,并且为了缩短产品开发周期,提高设计质量,这些设计有时是并行地进行。由于产品整体性能和结构的约束,产品各部分之间必然存在着很多约束关系。每个单位都将产品中属于自己设计的那部分的设计结果送入设计冲突检测工具,由它进行产品总体性能、制造、装配上的冲突检测。

由于产品各个部分之间在结构上、性能上存在着相互的联系,因此在产品某一部分的设计发生了更改时,必须及时将设计的更改情况通知与这个部分的设计有关的合作设计参与方,以便各个参与方及时调整自己的设计。

产品设计过程中会涉及大量的数据。为了设计的顺利进行,必须对产品全生命周期的数据进行有效的管理,保证数据的完整性和一致性。

### 4.2 安全控制

安全控制一直是计算机系统中很重要的问题,对互联网上的合作设计系统来说则显得尤为重要。互联网上的合作设计系统中的安全控制主要体现在:合作设计中成员身份的验证、成员权限的控制和数据的加密与解密。重点应该是保证有关产品全生命周期的设计数据在合作设计成员中安全、有效地传输和验证。

安全控制的主要技术有防火墙、用户身份验证系统、数字签名和数字水印等。

## 附录

## 知识资源 UDDI 注册信息填写表

## UDDI Registration Information Tables for An Intelligent Resource

表 49.4-2 账号注册信息 (Registration Information)

账号名称 Account Name	
账号密码 Account Password	
电子邮件 Your E-Mail	
本表管理员 Administrator's Name	先生/女士 Mr./Ms.      职务 Position
缺省语言 Default Language	中文 Chinese
公司/单位名称 Company/Unit Name	
电话号码 Your Phone Number	
传真 Your Fax Number	
详细地址 Detail Address	
国家 Country	中华人民共和国 P R China
邮政编码 Zip/Postal Code	

注：请在进入 UDDI 注册中心后修改您的初始账号密码。

Note: Please change your initial account password after you enter the UDDI Registry.

表 49.4-3 知识服务实体信息 (Knowledge Service Entity/BusinessEntity Information)

知识服务实体名称 Knowledge Service Entity Name	(中文) (English)
描述 Description	(中文) (English)

表 49.4-4 知识服务实体联系人信息 (Knowledge Service Entity Communication Information)

姓名 Person name	先生/女士 Mr./Ms.      职务 Position
描述 Description	(中文) (English)
电话 Phone Number	
传真 Fax Number	
电子邮件 E-Mail Address	
地址 Mail Address	(中文) (English)

注：如果需要填写的知识服务实体联系人信息超过一个，请您复印该表（并），填写并在姓名前加序号。

Note: If the number of person is more than one, please copy this table, fill it and put a number before the person name.

表 49.4-5 知识服务实体服务信息 (Knowledge Service Entity Service Information)

服务名称 service Name	(中文) (1) (English)
服务描述 Service Description	(中文) (English)

(续)

服务名称 service Name		(中文) (1) (English)
绑定信息 Binding templates/ information	入口地址 Access Point	
	类型① Kind	
	描述 Description	(中文) (English)

① 绑定信息的类型请选择: mailto/http/https/ftp/fax/phone/other。

Note: Please select the kind of binding templates/information within: mailto/http/https/ftp/fax/phone/other.

注: 如果需要填写的知识服务实体服务信息超过一个, 请您复印该表, 填写并服务名称前加序号。

Note: If the number of service information is more than one, please copy this table, fill it and put a number before the service name.

## 第5章 现代设计与制造网上合作研究中心及相关的资源

### 1 中心的创建与进展

#### 1.1 中心的创建

1995年底,国家自然科学基金委员会组织一个小组着手撰写“先进制造技术基础优先领域战略研究报告”。在一年多的准备过程中,参与撰写的人意识到除体制和经营管理两个层次的原因外,导致目前我国制造业困难的一个重要因素是企业缺乏开发有竞争力的产品的能力,因此提出了“产品设计是制造业的灵魂”的口号<sup>[1]</sup>。后来,寻求解决该问题途径的努力促成了创建网络合作设计组织来宣传互联网上的合作设计的理念,并且以实际的技术、应用、工具和资源等来支持互联网上的合作设计的实践活动。

1997年,“现代设计与产品研究开发网络—虚拟异地合作设计组织”(以下简称“合作设计组织”)宣布成立,这也是国内比较早的以促进互联网上的合作设计,支持产品的网络合作设计为目的的研究开发型网络。

2001年,教育部决定成立网上合作研究中心,在现代设计和制造领域依托西安交通大学和华中科技大学成立教育部现代设计与制造网上合作研究中心(Internet-Based Collaborative Research Center on Modern Design and Manufacturing of Ministry of Education of China, IBCDM,以下简称中心)。2002年



图 49-5-1 现代设计与制造网上合作研究中心网站主页

又增加了上海交通大学。目前中心在学术委员会和主任之下分为中心(西安)、中心(武汉)和中心(上海)三个部分。中心(西安)依托西安交通大学,侧重设计方面的研究;中心(武汉)依托华中科技大学,侧重制造方面的研究;中心(上海)依托上海交通大学,侧重将设计和制造方面的研究成果应用于企业中。三方通过网络联系作为一个整体开展工作。网站是中心对外的窗口,中心(西安)和中心(上海)的网址为 <http://www.chinamoderndesign.com>, 主页如图 49-5-1 所示,同时也是合作设计组织的网站地址,中心(武汉)的网址为 <http://www.amitnet.hust.edu.cn>。

#### 1.2 中心的进展

中心和合作设计组织成立后的头几年中,一直在摸索中前进。在许多热心的参与者的积极支持和密切配合下,中心和合作设计组织逐步取得了一些进展,可以归纳如下:

1) 有更多的拥有资源或潜在资源的单位将他们的主页链接到了中心的网站上,特别是许多还把他们的下属于单位的网站与中心的网站进行了分别的链接,这样客户在搜索信息和资源时就更为方便。更重要的进展是一些零部件供应商开始加盟,如中国轴承信息网、上海电机厂诺登轴瓦有限公司等,构成了设计信息来源和知识资源、知识获取资源的一个重要的方面。

2) 中心长期以来一直致力于基于互联网的设计知识和知识获取服务的研究和应用,目前能够提供基于网络的多种设计知识和知识获取服务。根据我们多年的工作,已有的资源单元(分布在全国各地)在互联网上提供知识和知识获取服务可以有四种形式:

##### (1) 远程程序调用和远程数据库服务

这是提供已有知识和知识获取服务(含数据库、知识库、数值分析程序等)在互联网上在线实时的知识服务的主要方式,在这种形式下,计算程序以及数据库已经可以通过网络来访问,在浏览器中找到相应的网址,填入需要的参数或者搜索关键字,可以进行计算或者查询数据库中的数据,然后将计算结果或者数据库查询结果显示在浏览器中供用户使用。包括基于网络的 300MW 汽轮机密封动力系数计算程序<sup>[2]</sup>, 转了一轴

承系统稳定性计算程序<sup>[42]</sup>和滚动轴承数据库<sup>[43]</sup>等。如果将相关的计算程序或者数据库采用 Web Service 的形式来访问,用户还可以将服务集成到自己的程序或者应用中,实现参数调用和返回过程的进一步集成。这也是远程程序调用和远程数据库服务的主要发展方向。

#### (2) 远程分析研究

对于需要人介入的服务,用户需要将数据或者图形传递到服务提供方,结果不能在当时返回,而是在双方约定的时间中返回。例如某些分析程序耗时较长,或者需要服务提供者进行某些离线的分析,这时用户不必在线等待,服务提供方完成分析后,即可将结果传递给用户。

#### (3) 远程实验

对于目前在技术上还不能依靠网络来调用的资源,由能提供服务的这一类单位(供应商)把他们的服务项目、技术范围、网上委托办法、合同样本等链接到网站上的网上服务栏目,用户通过网络了解服务提供方提供的实验服务的性质和范围,选定实验服务的名称,下载实验委托单,在与提供方就要求、付费、时间协商一致后,由委托方或提供方做好试件并进行实验。在接受委托后,服务方以离线的方式完成委托,将实验结果通过多种方式传递给用户。例如,西安交通大学润滑理论及轴承研究所已经把他们的所有可以对外支持产品开发的实验室资源按照上述要求在网上公布,其中包括国内少有的 200mm 单个流体动压轴承试验台、30000 (r/min) /50mm 轴颈直径/1000mm×2 跨转子试验台等。目前发展的方向是实现委托方可以在异地从浏览器上自己控制实验参数、看到实验的场景、听到实验的声音并立刻得到实验的结果。例如,该研究所目前提供的转子远程实验就可以达到上述要求<sup>[24]</sup>。

#### (4) 远程设计会议

在需要请求方与服务提供方详细讨论之后才能确定服务细节时,可以通过在电子白板上共同讨论修改图形、通过摄像头进行实物观察、通过话筒进行讨论以及联合运用前 3 种方式对不能决定的问题共同研究有关的信息,使请求方和提供方以及其他利益相关方最终取得一致意见。例如,上海交通大学生命质量与机械工程研究所与西安交通大学润滑理论及轴承研究所进行合作,由后者利用自己的网络资源在西安地区作为前者的代理,接受西安地区的外科医生,向生命质量与机械工程研究所订购定制型假体的业务。同时,润滑理论及轴承研究所向医生提供在西安与上海视频和音频讨论的条件,以保证客户和供应商双方能对定制型假体的技术要求进行充分讨论并签署合作合同。这是在网

上较早的能够使不同地理位置的医生和假肢制作单

位在网络上协商假肢定制的合作。

3) 开通了现代设计网上讨论栏目,在网上对现代设计这样非常重要但又缺乏经验的问题展开讨论并听取各方面对网站工作的意见。还链接了国内其他重要的设计、制造网站的论坛。

4) 组织了关于现代设计的网上教育。

## 2 中心的网上资源介绍

现代设计与制造网上合作研究中心网站上提供的网上资源从技术和服务的角度可以分为性能分析评估服务、支持设计的数据库服务、性能实验评估服务、可制造性评估服务、在运行产品状态评估服务、服务供应商的评估服务、虚拟仪器服务等。

### 2.1 性能分析评估服务

中心网站上提供的性能分析评估服务比较多,下面以滑动轴承性能计算网络服务为例进行介绍。

滑动轴承性能计算网络服务软件是西安交通大学润滑理论及轴承研究所(以下简称轴承所)利用多年积累的在滑动轴承性能计算方面的算法及经验而开发的一个网络服务应用软件。该程序可以计算的滑动轴承种类包括:径向轴承包括普通圆瓦、椭圆瓦、错位瓦、多油叶和可倾瓦;推力轴承包括面推力瓦和可倾瓦。

程序提供良好的输入和输出界面。输入参数和输出参数(包括许用判定参数)设计在同一界面上,方便用户输入参数,计算后直接可见输出参数,并判断输出参数是否合适,同时将计算结果保存到后台数据库。如果计算得出的参数不合适,用户就可以方便调整某个输入参数,重新计算,直到满足用户设计要求。需要时还可以生成结果报表。普通圆瓦滑动轴承的计算界面如图 49.5-2 所示。

输入参数为有量纲参数,针对不同类型的滑动轴承,包括:偏心率(直径间隙);各种轴承结构、几何参数(影响轴承使用特性的有关参数,描述与轴承图形示例一致);轴的转速、轴承受力的大小(径向力、轴向力)、力在轴承中的位置;进油温度、压力、油品参数。

输出参数为有量纲参数,针对不同类型的滑动轴承,包括:许用载荷(轴承比压)、最小油膜厚度、承载区最高工作温度、失稳转速、偏心距、油量、功耗、轴在轴承中的位置、最小油膜位置相对轴承剖分面的夹角、刚度系数、阻尼系数等。

每次计算后,可以选择生成包括计算类型、输入参数、输出参数在内的计算结果报表。多次计算时,程序提供了查询计算记录的功能,用户可以用输入参数和



图 49.5-2 普通圆柱滚动轴承性能计算页面  
输出参数以及计算时间等多种查询关键字进行复合查询，显示查询结果。

滚动轴承性能计算网络服务程序是基于 Web 服务技术开发的，其他系统可以通过 Web 引用的方式，方便的集成这些性能计算算法。图 49.5-3 为滚动轴承性能计算 Web 服务程序的实现过程。

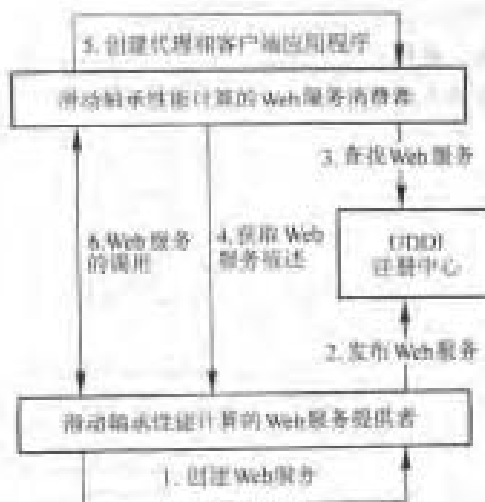


图 49.5-3 滚动轴承性能计算 Web 服务程序的实现过程

## 2.2 支持设计的数据库服务

在中心网站上提供的支持设计的数据库服务比较多，下面介绍基于 Internet 的滚动轴承数据库系统。

基于 Internet 的滚动轴承数据库系统采用了网络数据库技术、HTML 和动态服务器页面 (Active Server Page, ASP) 编程技术，神经网络及 CAD 技术，可以从 Internet 访问和使用，能够为用户提供滚动轴承设计数据库服务。系统主界面如图 49.5-4 所示。



图 49.5-4 基于 Internet 的滚动轴承数据库系统主界面

### 2.2.1 系统的主要功能

基于 Internet 的滚动轴承数据库系统的主要功能如下，功能结构图如图 49.5-5 所示：

- 1) 产品信息浏览，主要是滚动轴承产品基本知识的浏览，包括：滚动轴承的基本结构与分类，滚动轴承的代号，滚动轴承的选择，滚动轴承的额定载荷与寿命，滚动轴承的极限转数等。



2) 数据信息查询: 主要包括滚动轴承旧代号查询、新代号查询、模糊尺寸查询以及按轴承的适用工况条件查询等查询方式。

3) 图形信息交互: 可以实现滚动轴承的

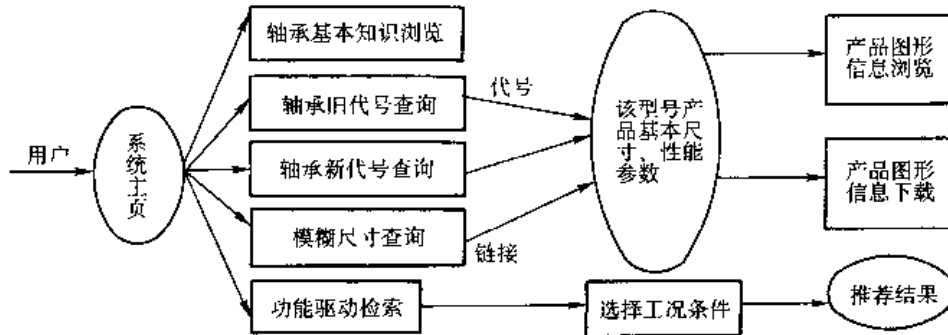


图 49.5-5 基于 Internet 的滚动轴承数据库系统功能结构图

### 2.2.2 系统的特点

1) 采用浏览器/服务器的实现模式。用户端不需任何设置。只要通过本机的浏览器, 即可以访问到远程数据库, 所需操作就是在网页输入自己的信息, 点击相关链接, 操作结果也可以网页形式显示给用户, 直观方便。

2) 功能完善。系统实现了产品信息浏览, 为用户提供了大量关于产品的基本知识。产品信息检索为用户提供了多种检索方式, 有代号检索, 模糊尺寸检索和功能驱动检索。产品数据信息交互为用户提供了产品数据信息的浏览与下载。产品图形信息交互为用户提供产品图形信息的在线浏览和下载, 用户可对查询到的产品的图形信息进行在线浏览, 还可以把图形信息下载到本地机上, 直接在设计中使用。

### 2.3 性能实验评估服务

轴承所在中心的网站上提供了性能实验评估服务, 可以在轴承性能、转子系统、铁谱技术、电磁轴承和远程状态检测等多个方面提供性能实验评估服务。委托方根据所需要的服务填写委托书, 并通过协调, 与研究所达成委托协议, 即可开展性能实验评估服务。目前可以承接委托实验及加工项目如下所示。

1)  $\phi 200$  径向滑动轴承性能试验台: 主要参数为试验轴承内径  $D_{\max} = 200\text{mm}$ , 转速  $n_{\max} = 10,000\text{r/min}$ , 载荷 = 4T, 试验内容为径向滑动轴承的动、静特性测试。

2)  $\phi 100$  径向滑动轴承性能试验台: 主要参数为试验轴承内径  $D_{\max} = 100\text{mm}$ , 转速  $n_{\max} = 20,000\text{r/min}$ , 载荷 = 1T, 试验内容为径向滑动轴承的动、静特性测试。

Solidworks 三维图形的在线浏览, 以及下载滚动轴承的 Unigraphics 三维图形, 所有滚动轴承的三维图形可以直接在产品设计中, 不需重新绘制。

3) 转子—轴承系统试验台 1: 主要参数, 轴颈  $\phi 50$ , 跨距为单跨 1000mm, 双跨  $2 \times 1000\text{mm}$ , 转速为  $n_{\max} = 30,000\text{r/min}$ , 试验内容为转子动力学试验, 振动测试与信号分析, 轴承转子系统的性能测试与参数识别, 故障模拟诊断等。

4) 转子—轴承系统试验台 2: 主要参数, 轴颈  $\phi 15$ , 跨距为单跨 300mm, 双跨  $2 \times 300\text{mm}$ , 平行轴, 转速为  $n_{\max} = 10,000\text{r/min}$ , 试验内容为转子动力学试验, 振动测试与信号分析, 参数识别, 故障模拟诊断, 齿轮耦合对系统的影响等等。

5) 铁谱分析: 铁谱技术是利用磁场作用, 将机器润滑油样中的铁质磨屑粒子和粘附物微粒分离出来, 并分析其形态、尺寸、数量、粒度分布等情况, 从而得到磨损过程的有关信息, 利用铁谱技术, 可以对机械设备运行中的磨损状态进行诊断、监测, 从而提高系统的可靠性和安全性。铁谱技术的重要工具之一就是铁谱仪。轴承所可以提供基于铁谱仪的油液磨粒分析服务。

6) 电磁悬浮轴承性能分析设计制造、试验服务: 电磁悬浮轴承利用电磁力使高速旋转的转子悬浮而与轴承没有任何接触, 电磁悬浮轴承无接触、无润滑、无磨损、功耗低, 工作寿命长, 能够允许转子高速运转, 其转速仅受转子所选材料强度的限制。电磁悬浮轴承主要应用在真空及超净室技术、高精度机床、透平机械和离心机等方面。已完成项目包括电磁悬浮轴承支承的  $\text{CO}_2$  激光风泵, 电磁悬浮轴承支承的导弹质量、质心、转动惯量测试系统, 电磁悬浮轴承支承的 100 万方涡轮膨胀机, 电磁悬浮轴承支承的高速磨床主轴等。轴承所能够为电磁悬浮轴承的用户提供技术设计、技术咨询和委托制造等服务。

7) 系统远程状态监测及诊断技术服务: 轴承所在大型旋转机械机组状态监测故障诊断领域有数十年的

经验,取得了大量科研成果,积累了丰富的分析故障和解决问题的经验。研究所针对炼油、化工、电力企业的主要动力设备开发了RB21 机组状态监测及故障诊断系统,能够为用户提供远程状态监测及诊断技术服务。

## 2.4 服务供应商的评估服务

中心网站上提供的 AMT 供应商评价系统是西安交通大学先进制造实验室开发的服务供应商评估应用系统,它分为供应商信息管理、指标体系管理及供应商评价三大模块,下面分别予以介绍。

### 2.4.1 供应商信息管理

在供应商信息模块中可以添加、删除、修改各供应商的信息,界面如图 49-5-6 所示。



图 49-5-6 AMT 供应商评价系统供应商信息管理界面

### 2.4.2 指标体系管理

根据需要,在指标体系管理中,如图 49-5-7 所示,给不同的指标体系设置了不同的评价指标。目前有生产性组织评价指标体系、创新性组织评价指标体系和知识服务供应商评价指标体系三种指标体系,其中生产性组织评价指标体系应用于非创新性产品(或服务)供应商的评价与选择(主要针对一般生产企业),创新性组织评价指标体系应用于创新性产品供应商的评价与选择(主要针对愿意为特定需求研发新产品的企业、科研院所和高校等),知识服务供应商评价指标体系应用于知识或知识获取服务供应商的评价与选择(主要针对愿意以知识或知识获取服务参与产品研发的企业、科研院所和高校等)。

用户可以根据自己的需要创建、编辑和删除指标体系。在每种指标体系中,包含有定性和定量的指标,用户也可以根据需要创建、编辑和删除相应指标体系中的指标。对于知识服务供应商评价指标体系就有 14

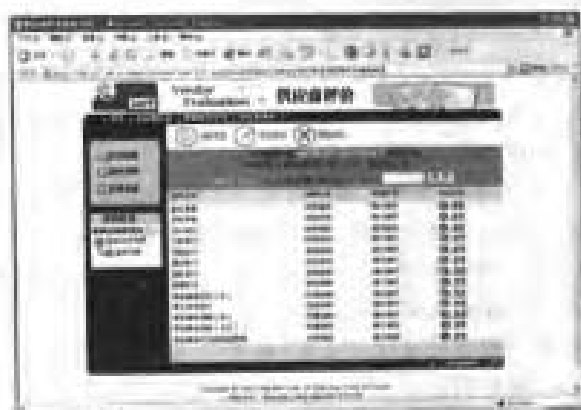


图 49-5-7 AMT 供应商评价系统指标体系管理界面  
个不同的评价指标,分别如下:

- 1) 服务质量,指标类型为定性指标,指标性质为越大越好。
- 2) 研发声誉,指标类型为定性指标,指标性质为越大越好。
- 3) 设计能力,指标类型为定性指标,指标性质为越大越好。
- 4) 分析能力,指标类型为定性指标,指标性质为越大越好。
- 5) 试验能力,指标类型为定性指标,指标性质为越大越好。
- 6) 通讯能力,指标类型为定性指标,指标性质为越大越好。
- 7) 团队实力,指标类型为定性指标,指标性质为越大越好。
- 8) 管理水平,指标类型为定性指标,指标性质为越大越好。
- 9) 项目研发历史(年),指标类型为定量指标,指标性质为越大越好。
- 10) 项目研发现状,指标类型为定性指标,指标性质为越大越好。
- 11) 项目研发周期(年),指标类型为定量指标,指标性质为越小越好。
- 12) 项目研发预算(万元),指标类型为定量指标,指标性质为越小越好。
- 13) 项目研发技术指标的预期,指标类型为定性指标,指标性质为越大越好。
- 14) 对需求满足的程度的预期,指标类型为定性指标,指标性质为越大越好。

### 2.4.3 供应商评价

供应商评价的步骤为:

- 1) 选择评价指标体系,根据不同的评估目的,选

择现有的评估指标体系, 或者根据需要添加新的评估指标体系和体系中的指标。

2) 可行性评价, 先可以在选定的评估体系中确定至少三个关键指标, 根据这些关键指标在供应商信息数据表中选择出合适的供应商。

3) 帕累托最优过滤, 对于一个供应商而言, 如果无法找到一个替代方案(指潜在的其他供应商或者潜在的其他供应商的一个线性组合)可以改进该供应商的某项指标而不使其他任何指标受损, 那么这个供应商就称为帕累托最优供应商(这里的改进的含义是, 对于越大越好的指标, “改进”意味着指标值变大; 对于越小越好的指标, “改进”则意味着指标值变小, “受损”则是相反)。如果一个供应商不是帕累托最优的, 那么就可以找到一个更满意的新方案完全替代这个供应商, 如果一个供应商是帕累托最优的, 这个供应商不可能被完全替代, 并令其不可替代性系数为“1”。



图 49-5-8 AMT 供应商评价系统指标偏好设置



图 49-5-9 AMT 供应商评价系统知识服务供应商评价结果

4) 满意度评价, 根据层次分析法 (Analytic Hierarchy Process, AHP), 对越小越好的指标整体偏好, 越大越好的指标整体偏好, 研发能力类指标的偏好, 项目特定指标类指标的偏好进行设置, 如图 49-5-8 所示。设置指标偏好后, 对所有能够提供服务的供应商进行满意度评价, 从而选出最适合的供应商, 评估结果如图 49-5-9 所示。

## 2.5 虚拟仪器服务

### 2.5.1 背景及意义

随着计算机和网络技术的发展, 在互联网上提供虚拟仪器服务成为可能, 这也是网上合作设计知识资源服务的一种重要方式。在这种服务中可以做到对远程实验设备和仪器的操作、控制和测试, 并对实验的数据进行实时的处理和分析。远程诊断中心在中心网站上提供了基于小型多功能转于试验台的远程虚拟仪器实验服务的资源, 开展了网络远程教育和远程虚拟仪器服务的研究。

### 2.5.2 功能介绍

远程虚拟仪器实验系统采用 BridgeView 构造本地的虚拟仪器商控系统, 不仅可以方便灵活地采集、分析与处理各信号, 而且还能提供良好的仪器操作界面以便监控和干预远程用户的试验过程。异地用户可以通过 Web 浏览器连接到远程虚拟仪器实验网页, 自主控制实验参数, 并可以即时看到实验的场景, 听到实验的声音并立刻得到实验的结果。基于 Internet 的远程虚拟仪器实验系统总体结构图如图

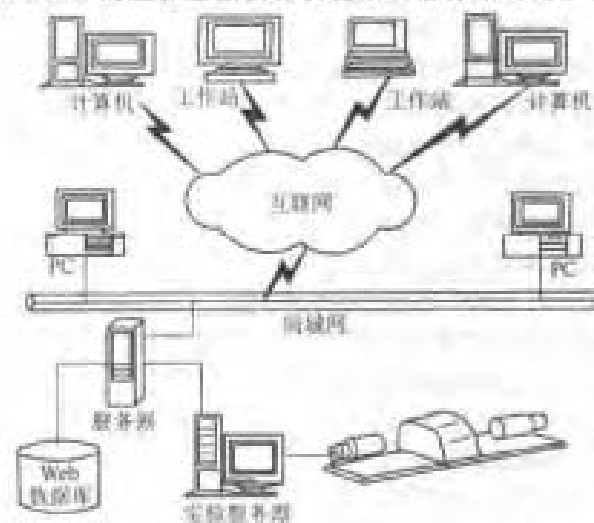


图 49-5-10 基于 Internet 的远程虚拟仪器实验系统总体结构图

49-5-10 所示。服务器端和客户端主界面如图 49-5-11、49-5-12 所示。

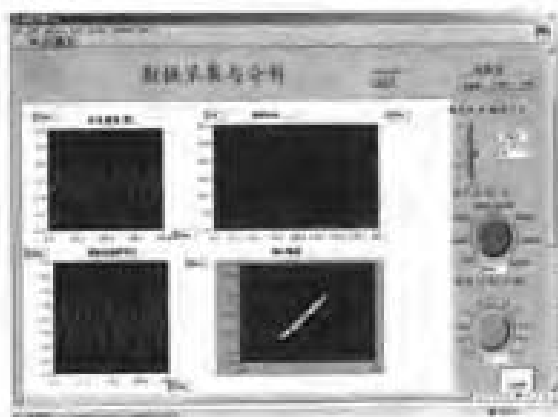


图 49-5-11 基于 Internet 的远程虚拟仪器实验系统服务器端界面

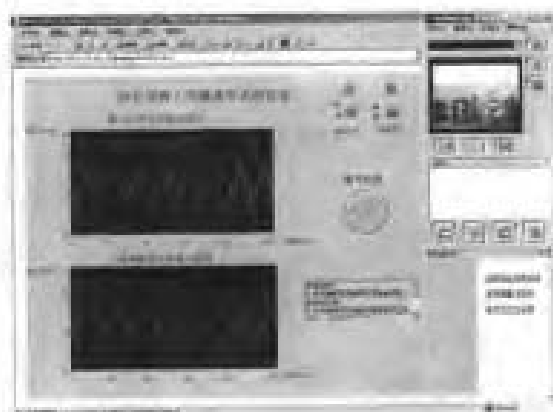


图 49-5-12 基于 Internet 的远程虚拟仪器实验系统客户端界面

### 3 中心资源的使用导航

在现代设计与制造网上合作研究中心网站上提供了许多可以提供知识服务的智力资源,为了更好的方便用户使用中心提供的资源以及解决如何组织远程会议的问题,这里就如何使用智力资源的统一发布、发现和集成(UDDI)注册中心和组织远程会议做一简单介绍。

#### 3.1 如何使用智力资源的统一发布、发现和集成(UDDI)

中心采用的 UDDI 注册中心为 Systemet 公司提供的 WASP UDDI 系统。WASP UDDI 是目前成熟的、安

全的和平台独立的 UDDI 注册服务系统,主要用于专有 UDDI 注册中心的建立和注册信息的管理。中心网站以 WASP UDDI 系统为基础建立的用于互联网上的合作产品设计的 UDDI 注册中心的主界面如图 49-5-13。



图 49-5-13 中心网站提供的支持互联网上的合作产品设计的 UDDI 注册中心的主界面

##### 3.1.1 用户注册

如果希望在 UDDI 注册中心发布和搜索有关知识服务实体和服务信息,就必须先进行用户注册。UDDI 注册中心提供了用户注册功能,在主页面进入注册页面,填写相关信息即可完成。

##### 3.1.2 发布知识服务实体和所提供的服务的信息

用户注册后,在主页面进入发布页面,即可填写商业实体信息(businessEntity)、服务信息(businessService)、绑定信息(bindingTemplate)、服务调用规范(tModel)等信息,图 49-5-14 显示了发布知识服务实体和所提供的服务的信息的界面。

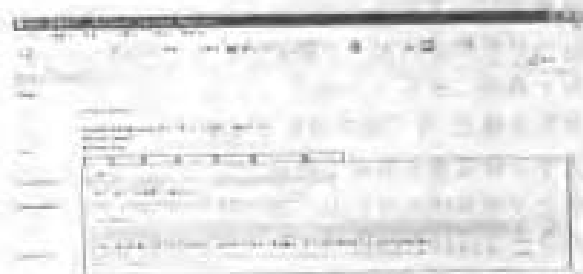


图 49-5-14 UDDI 中心提供的发布知识服务实体和所提供的服务的信息的界面

### 3.1.3 搜索知识服务实体和所提供的服务的信息

知识服务实体发布信息以后,企业等其他实体就可以通过 UDDI 注册中心提供的搜索功能,如图 49-5-15 所示,使用 Find business, Find service, Find binding, Find tModel 等模块,进行查询,找到相关服务的交互及接口等信息,就可以调用服务单元提供的服务了。



图 49-5-15 UDDI 中心提供的搜索知识服务实体和所提供的服务信息的界面

## 3.2 如何组织远程会议

互联网上的合作设计需要提供各种网络多媒体实时交互平台和图形实时互操作平台来支持远程会议的实现。这里以远程假体异地合作设计为例来说明如何组织远程会议<sup>[2]</sup>。

### 3.2.1 远程会议实现背景

上海交通大学生命质量与机械工程研究所基于生物摩擦学对人工关节设计的相关技术进行了研究,研制成功计算机辅助人工关节置换临床工程软、硬件,并且与上海第二医科大学第九人民医院进行临床实践,手术取得成功。为了实现该科研成果的产业化,上海交通大学生命质量与机械工程研究所与西安交通大学润滑理论及轴承研究所协议,通过西安交通大学网络设施开展远程假体合作设计业务。后者利用自己的网络资源在西安地区作为前者的代理,接受西安地区的外科医生向生命质量与机械工程研究所订购定制假体的业务,同时,润滑理论及轴承研究所向医生提供在西安与上海视频和音频讨论的条件,以保证客户和供货商双方能对定制假体的技术要求进行充分讨论并签署合作合同。

### 3.2.2 远程假体异地合作设计的业务流程

远程假体异地合作设计的业务流程如图 49-5-16 所示。



图 49-5-16 远程假体异地合作设计的业务流程

### 3.2.3 远程会议实现的环境

基于上面对系统流程的设计和 demand 分析,建立如下环境:

#### (1) 硬件环境

实现该系统除了两端计算机配置高清晰度摄像头以外,对网络传输带宽的也有一定要求,以保障多媒体数据的实时交互,各种图像数据的传输。实际会议中,在远程会议两端计算机分别通过局域网连接到校园网,西安交通大学和上海交通大学分别建有千兆校园网,并且分别为教育科研网在西北和华东(南)的地区中心,与北京清华大学教育科研网中心相连。

### 2) 软件环境

软件环境使用微软公司的 NetMeeting 3.0 作为视频会议软件。NetMeeting 是一个综合性的网络通讯软件,它提供了强大的网络实时交互功能,包括视频传输、音频传输、在线文字交互、白板交互、远程程序控制和远程桌面共享等功能。NetMeeting 提供基于标准的多点数据会议,用户可与两个或更多的与会者进行实时的合作和信息共享。通过视频捕获卡和摄像头,NetMeeting 的用户可在会议过程中发送和接收视频图像,从而进行面对面的交流。

软件环境使用 Materialise 公司的 Magic Communicator 作为三维图形实时交互浏览的工具。Magic Communicator 是在 Internet 或 LAN 上讨论三维

CAD 图形的软件。在实时在线网络会议中,用户可以共享他们的三维 CAD 文件及标注、尺寸测量等,由于灵活使用了 Internet 协议和数据压缩技术,即使在使用防火墙的情况下,消息传送时间也在 2s 之内,交互操作基本达到了实时的效果。此外它还通过口令和特定的通讯服务器提供安全保障性能。Magic Communicator 提供多种 CAD 文件格式的浏览支持,包括 STL, DXF 3D 曲面, SolidWorks, SolidEdge, IGES, VDA, Catia, Unigraphics 和 Parasolid 等。

系统交互实时环境如图 49-5-17 所示。通过实际运行,上述系统能够满足假体远程设计中交互讨论的要求。音频信号传递实时、清晰,视频信号传递连贯,交互操作的实时性可以满足网络讨论的要求。



图 49-5-17 远程会议系统交互实时环境

## 4 中心的发展方向

几年来的实践说明,实现互联网上的合作设计并非易事,有三个层次的问题需要解决。

首先的层次是观念上的问题。现在大家都讲竞争,竞争是优胜劣汰,强强联合优于单干,所以合作在竞争中占有极其重要的位置。如果把在网上的就某一个任务的合作看成是组织一个虚拟公司,虚拟公司讲究合作的愿望,而使愿望成为现实的重要一环,是参与各方要接受利益在参与者中合理分配这一原则,而且合作各方也要承认合作有利于在竞争中的生存和发展,有利于比单干获得更大的市场份额和利益。

第二个层次是管理上的问题。管理不好的合作当然还不如单干,愿望要靠组织和游戏规则(法律)来保证,这里面有两个具体问题:一是参与网上合作的各方所提供的知识资源的知识产权问题;二是利益在参与网上项目组织之间的合理分配问题。因此合作设计组织的“管理小组”,起草了一个文件(附在本章后),供网上组织合作时参考,当然可以根据具体情况增减各

方的权利和义务。

第三个层次是技术上的问题,包括网上运行的规范和资源建设协调等。

最难解决的是前两个层次的问题,需要在良好的合作和成功之中逐步建立信心,预料要有一个较长的磨合过程,信心是不能依赖行政手段建立的,所以市场驱动是唯一的选择。

至于技术层次的问题,网络速度是一个较大的问题,各种工具所依赖的通讯协议不同,也给设计的合作带来许多困难。相信这些问题能够在可预见的将来逐步得到解决。

## 附录

### 虚拟研究开发中心合同 (供参考)

#### 总则

合作方式  
经营管理机构

知识产权  
利润分配和亏损分担  
解散和清算  
附则

本合同供组织或参加虚拟研究开发中心时参考使用。

## 总则

第一条 根据《中华人民共和国民法通则》、《中华人民共和国合伙企业法》和《中华人民共和国公司法》，\_\_\_\_\_(以下简称甲方)、\_\_\_\_\_(以下简称乙方)与\_\_\_\_\_(以下简称丙方)于\_\_\_\_\_年\_\_\_\_\_月\_\_\_\_\_日签订建立本虚拟研究开发中心(以下简称虚拟中心)合同，并制订本章程。

第二条 甲方的名称为\_\_\_\_\_，法定地址为\_\_\_\_\_。乙方的名称为\_\_\_\_\_，法定地址为\_\_\_\_\_。丙方的名称为\_\_\_\_\_，法定地址为\_\_\_\_\_。

第三条 虚拟中心的名称为\_\_\_\_\_，外文名称为\_\_\_\_\_。

第四条 虚拟中心是非法人的营利性经济组织。

第五条 虚拟中心的宗旨是：利用计算机网络等电子手段，不同研究单位针对某具体研究开发项目，在短期内迅速建立相互信任的合作关系，捕捉市场机遇，利用自己具有的核心能力(优势)获取令合作各方满意的经济利益。

第六条 合作各方在筹建虚拟中心和相互合作过程中订立各项书面合同或电子合同，应当遵循自愿、平等、公平、诚实、信用的原则。

第七条 虚拟中心及合作各方的财产和合法权益受法律保护。

第八条 虚拟中心的建立和解散均应以书面形式或电子手段公告。

## 合作方式

第九条 合作各方的法律地位不因虚拟中心的建立而改变。

第十条 虚拟中心无权干涉合作各方的，与该合作项目无关的其他经营问题。

第十一条 虚拟中心的合作各方在进行合作之前，共同或委托一方进行可行性研究，预算的全部研究费用由合作各方共同负担，具体比例：甲方\_\_\_\_\_%，乙方\_\_\_\_\_%，丙方\_\_\_\_\_%。

第十二条 合作各方应当按照合同规定的期限缴清各自负担的费用，形成合作基金，合作过程中需要增

加的费用按第十一条规定比例，由合作各方共同负担。

第十三条 新的合作方加入虚拟中心时，应当经过董事会批准，并依法订立书面合同或电子合同，协商调整各自负担费用的比例。

第十四条 新合作方在订立合同时，原合作各方应当向新合作方告知虚拟中心研究开发工作总体规划、实际进展、财务状况以及其他应当由其知道的内容。

第十五条 新合作方与原合作方享有同等权利，承担同等责任。若合同另有规定的，以其约定为准。

第十六条 有下列情形之一的，合作一方可以退出虚拟中心：

(一) 发生合作方难以继续参加合作的事由，如破产等，经过董事会讨论同意。

(二) 其他合作方严重违反合同约定的义务或出现严重的信用危机。

(三) 合同约定的其他退出事由出现时。

第十七条 合作方退出时，其负担的研究开发费用，应根据工作的实际进展和合同的规定予以退还。因其退出而给虚拟中心造成损失的，应由其负责赔偿。合作各方另有约定的，以其约定为准。

第十八条 合作方退出时，仍应当按照合同规定对虚拟中心和其他各合作方的商业秘密承担保密义务，若有违反，按合同的规定承担相应的民事和法律责任。

第十九条 合作方的加入和退出，均应以实际发生日为其生效日。

## 经营管理机构

第二十条 虚拟中心设董事会，董事会为最高权力机构。董事会可以根据合同约定在董事会内部分工或委聘专人处理日常事物。

第二十一条 虚拟中心的董事会，由每个合作方各委派一名董事组成，各董事对于虚拟中心的经营管理享有同等的权利(除非合同事先有专门约定)。

第二十二条 虚拟中心的董事会通过计算机网络电子会议或电话等远程通讯方式开展工作。

第二十三条 虚拟中心的董事会职权如下：

(一) 修改虚拟中心的合同。

(二) 决定和批准合作项目的研究方向与工作计划。

(三) 批准费用分摊比例的变更及利润分配方案。

(四) 管理合作基金和审批费用的预算；批准增加研究开发费用。

(五) 批准新的合作方的加入，以及原有合作方的退出。

(六) 协调各合作方的合作关系。

(七) 评估合作各方用于合作的技术的先进性。

(八) 决定虚拟中心的解散，并主持期满时的清算

工作。

(九) 对外处理与虚拟中心相关的事宜。

(十) 任免处理日常事物人选。

(十一) 其他应由董事会决定的重大事宜。

第二十四条 关于虚拟中心任何提议,须经董事会三分之二以上多数通过方能生效。

第二十五条 合作各方可以对其他各合作方提出异议。提出异议时,各方应暂停该项目的执行,进行协调。若产生争议,可由董事会作出最终决定。

## 知识产权

第二十六条 合作方使用其他合作方所有或持有的专利技术的,按照《中华人民共和国专利法》和《中华人民共和国专利法实施细则》的规定,必须与其他合作方订立书面实施许可合同。被许可使用专有技术的合作方无权允许合同规定以外的任何单位和个人实施该专利。

第二十七条 合作各方均应当对自谈判开始至合作结束阶段交换和新产生的商业秘密(包括商业信息和专有技术)承担保密义务。保密期限另有约定的,以约定为准。

第二十八条 承担保密义务的人员,不得利用第二十七条所规定的商业秘密自营或为所在单位牟取利益。

第二十九条 合作研究开发中涉及合作各方的工程图纸、产品定义、数据模型、商务报告、计算机软件的,按照《中华人民共和国著作权法》和《计算机软件保护条例》处理。

第三十条 合作各方在合作研究开发中形成的所有智力成果,为合作各方共有。合作各方另有约定的,以约定为准。

第三十一条 前款所规定的智力成果可以被分割使用的,合作方对各自研究开发的部分可以单独享有所有权,但其所有权的行使不得构成对该智力成果其他部分的侵害。前款所规定的智力成果不能被分割使用的,该智力成果由合作各方共有。合作各方共同协商,一致行使其权利。若无法协商,也无正当理由,任何一方不得阻止他方行使转让权以外的其他权利,但所得收益应合理分配给其他合作各方。

第三十二条 合作各方在合作研究开发中形成的发明创造,申请专利的权利属于对该发明创造的实质

性特点,做出创造性贡献的合作方。申请被批准后,专利权归申请的单位所有或持有。享有专利申请权的合作各方必须协商一致,才能行使专利申请权。

## 利润分配和亏损分担

第三十三条 虚拟中心的合作各方,负有无限责任。董事会根据合作各方用于合作的技术评估及各自实际费用分担额,确定利润分配比例和亏损分担比例。

第三十四条 虚拟中心应依法缴纳所得税。

第三十五条 在虚拟中心存续期间,每个研究开发项目分配利润一次。项目为一次赢利的,董事会应在盈利后的一个月內公布利润分配方案和各方应得利润额;项目为多次盈利的,董事会在每次盈利后一个月內公布利润分配方案和各方应得利润额。

## 解散和清算

第三十六条 在有以下情形之一时,虚拟中心应当解散:

(一) 完成合同约定的研究开发项目。

(二) 发生研究开发项目难以继续事由,经董事会同意。

(三) 已经确定项目的研究开发不能达到预期的经济利益。

(四) 董事会决定终止合作研究开发,解散虚拟中心。

(五) 出现法律、法规规定的虚拟中心解散的其他原因。

第三十七条 虚拟中心解散时,清算组由董事会担任。清算组在清算期间按顺序执行下列事务:

(一) 清理合作基金。

(二) 处理与清算有关的虚拟中心未了结的事务。

(三) 清偿虚拟中心存续期间的各项费用。

(四) 按合作各方实际的费用分担比例返还所缴金额。

## 附则

第三十八条 本章程的修改和解释权归董事会所有。

甲方:(签章)

乙方:(签章)

丙方:(签章)



## 第 6 章 互联网上的合作设计实例

现代设计一直强调,不论是已有知识的使用,还是新知识的获取,都强烈地依赖具有 6 个不可分要素(人、资金、设备、经验、管理、技术)的资源。而通过什么模式来将这 6 个要素集成起来,学术界和产业界的许多人士提出了很多解决方案,例如产品数据管理(PDM),工作流管理系统(WFMS),企业资源规划(Enterprise Resource Planning, ERP),供应链管理(Supply Chain Management, SCM),协同产品商务(Collaborative Product Commerce, CPC)等。而知识服务的理论则是从实现现代产品设计的最基本要素—知识,以及知识的外在表现形式—资源,以及资源运作和服务的统一体—资源单元和设计实体出发提出的,实例说明在我国可以应用知识服务的理念来加快制造业的产品创新。

为了以实例说明如何利用分布式智力资源来开展基于互联网的合作产品创新设计,教育部现代设计及制造网上合作研究中心(IECDM)以涡轮膨胀机主动控制电磁悬浮转子系统的改型设计为对象(应某涡轮膨胀机制造企业—Turbo-Expander Co. Ltd, TEC 的请求),将国内有关最优智力资源集成起来,通过 Internet 完成一个完整的设计、分析和制造准备过程。合作单位包括:

- 1) 西安交通大学润滑理论及轴承研究所(Theory of Lubrication and Bearing Institute, TLBI)
- 2) 西安交通大学先进制造技术实验室(AMT laboratory, AMT)
- 3) 清华大学机械工程系(Department of Mechanical Engineering, DME)
- 4) 华中科技大学工业工程与制造系统工程系(Department of Industrial & Manufacturing Systems, DIMS)
- 5) 华中科技大学国家 CAD 支撑软件工程研究中心(National Engineering Research Center-CAD Software, NERC-CAD)

### 1 项目背景

项目的背景是 TEC 考虑到传统的采用动压滑动轴承支承的涡轮膨胀机逐渐不能满足市场竞争的需要,其了解到国外同类产品可以采用主动电磁轴承进行支承,所以它想重新设计涡轮膨胀机的支承系统,采用主动电磁轴承进行支承。TEC 的传统的产品设计主要在

企业内部完成,如果还是由企业自己学习、消化和掌握主动电磁轴承的设计方法和技术,效率较低,成本较高,企业很难尽快推出新产品,所以 TEC 希望能够通过网络合作设计来完成涡轮膨胀机采用主动电磁轴承的改型设计。

IECDM 提供了用于产品设计知识资源服务发布和发现的 UDDI 注册中心,其中已经有许多知识服务资源在其中注册。TEC 搜索 UDDI 注册中心,并评估

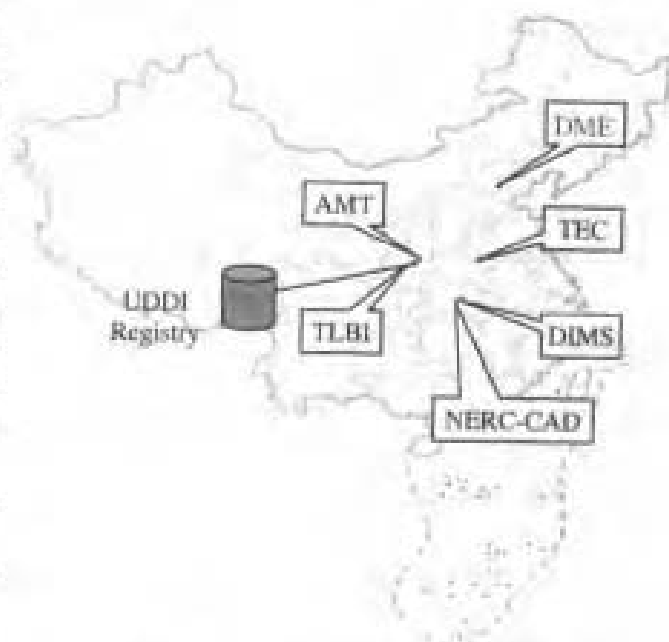


图 45.6-1 涡轮膨胀机采用主动电磁轴承

系统的合作创新设计的主要过程

1. TEC 搜索 UDDI 注册中心,寻找主动电磁轴承设计知识服务提供者
2. TEC 采用 AMT 提供的评估系统评估知识服务提供者,并选择 TLBI 为合作伙伴
3. TLBI 完成主动电磁轴承的结构设计
4. TLBI 完成轴承转子系统的动力学分析
5. TLBI 完成除磁设计
6. TLBI 搜索 UDDI,评估并选择 DME 完成可制造性评估,返回设计改进意见
7. TLBI 搜索 UDDI,评估并选择 DIMS 完成可加工性评估,返回设计改进意见
8. TLBI 搜索 UDDI,评估并选择 NERC-CAD 完成可装配性评估,返回设计改进意见

了相应的几个可以提供主动电磁轴承设计服务的设计服务提供商,最后选择 TLBI 来完成该项目。TLBI 设计了主动电磁轴承系统,包括控制器、功率放大器、执行器和传感器的设计,同时解决了轴承转子系统无法满足临界转速要求和掉电保护等问题。由于主动电磁轴承系统需要改变相关的涡轮膨胀机的结构,TLBI 通过 UDDI 注册中心,寻找并评估选择 DME、DIMS 和 NERC-CAD 完成相关结构的可铸造性,可加工性和可装配性的评估,并根据它们对设计提出的改进意见来优化设计。最终各单位通力合作,在短时间和低成本情况下很好的解决了相关的问题,完成全部的设计,并成功进行台架试验。图 49.6-1 为该改型设计项目的主要实现过程。

## 2 涡轮膨胀机采用动压滑动轴承支承的缺点

涡轮膨胀机主要在冶金、石化、化肥、轻纺、建材、能源等工业部门用于空气分离和气体液化,例如用于油气田制冷装置处理天然气等。目前空气分离行业采用的主要工艺是将空气、天然气等液化,利用不同成分液化温度的不同,分离出液氧、液氮或其他有用成分。因此,制冷设备是整个系统的关键,而涡轮膨胀机又是制冷设备中最重要的设备。对涡轮膨胀机的要求是轴承功耗小、长期运行可靠性高、易于实时监测、易于维护、成本低等。

TEC 的某型涡轮膨胀机组为双叶轮结构,工作转速为 30000~40000r/min。原系统采用动压滑动轴承支承,轴承功耗为 20kW,功耗大,效率低;润滑子系统体积庞大(比涡轮膨胀机主机大数倍);润滑油泄漏导致工质和环境污染。图 49.6-2 显示了采用动压滑动轴承支承的涡轮膨胀机的结构简图。

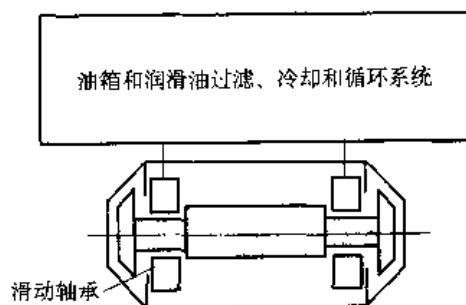


图 49.6-2 采用动压滑动轴承支承的涡轮膨胀机

## 3 涡轮膨胀机采用主动电磁轴承支承的优点

在磁悬浮领域中,应用最广泛的就是主动电磁轴承,它是利用电磁力将转子无机械接触地悬浮起来的

一种新型支承装置,是集转子动力学、电磁学、电子技术、控制理论以及计算机科学于一体的最具代表性的机电一体化产品。与传统机械轴承相比,电磁轴承具有无磨损、功耗小、不需润滑等特点;在一定范围内,电磁力可以通过控制系统予以调节,即支承刚度阻尼可调,因而可对转子实施主动控制、进行不平衡补偿等,有利于提高转子的动态性能,是目前可投入实用的主动控制型支承装置<sup>[26]</sup>。

与普通轴承相比,电磁轴承采用受控电磁力将轴旋浮在轴承中间,因而无接触、无磨损,省去了润滑系统,而相应增加了一套控制系统。鉴于电子元器件的寿命较机械系统长,因此,电磁轴承的可靠性及使用寿命优于普通轴承,同时系统功耗小,效率高。以该涡轮膨胀机组为例,电磁轴承的功耗小于 2kW,不及普通轴承的十分之一。电磁轴承长期使用成本要远低于普通轴承。在功率较大的场合,电磁轴承的一次投资成本也低于普通轴承。

涡轮膨胀制氧机机械结构简图如图 49.6-3 所示,左端为膨胀端,右端为压缩端。推力轴承与推力盘间的间隙由推力轴承座的尺寸确定,径向轴承与径向动环间的间隙由自身尺寸决定。

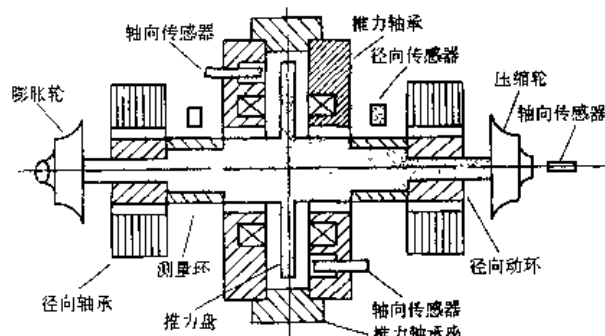


图 49.6-3 膨胀机机械结构简图

主动式电磁型磁轴承主要由四部分构成:传感器、控制器、功率放大器和执行器。其工作原理是:传感器检测出转子偏离参考点的位移,控制器则将这—位移信号转换成相应的控制信号,由功率放大器转换成一定的控制电流,控制执行器中磁铁吸力的大小,从而使转子恢复到原来设定的参考位置。

## 4 互联网上的合作设计过程

### 4.1 知识资源注册

基于网络的产品创新的前提条件是知识服务提供商将自己的服务发布到全球范围,同时知识服务请求者能够方便、准确地发现相关的合适的知识服务资源单元。目前得到广泛支持的统一描述、发现和集成(UDDI)协议提供了知识服务供应者和知识服务消费

者之间相互发现服务和需求的能力。这里所指的“服务”，可以认为是由知识服务提供商发布在网络上的完成特定知识服务的在线或离线应用服务，其他知识服务需求者或则应用软件能够通过 Internet 来访问、了解并使用这些网络上的服务。IBCDM 提供的用于合作产品设计知识服务供应商信息注册和管理的 UDDI 注册中心，各相关单位的智力资源单元提供的知识服务实体和实体提供的服务的结构化信息通过网络注册到该 UDDI 注册中心。

#### 4.2 搜索设计资源单元并评估

TEC 通过浏览器浏览 IBCDM 提供的基于统一描述、发现和集成协议的 UDDI 注册中心，以主动电磁轴承设计为关键字搜索，得知有数个单位可以提供主动电磁轴承的设计服务（以设计单位 A-G 代替实际单位名称）。图 49.6-4 显示了在 UDDI 注册中心填写搜索关键字和搜索所得到的知识服务实体。

为了在知识服务供应商中进行比较和选择，TEC 以设计服务评估为关键字搜索 UDDI 注册中心，发现西安交通大学管理学院先进制造技术实验室（AMT）提供的知识服务供应商评价系统，可以提供知识服务评估的功能；网络调用该程序，经过综合分析，选择 TLBI 作为合作者。



图 49.6-4 UDDI 注册中心搜索知识服务实体

#### 4.3 初步组成虚拟设计联盟

TEC 作为第一层设计实体用自有智力资源完成涡轮压缩机的前通流部分设计并提出改型的约束；TLBI 作为第二层设计实体设计电磁轴承，决定结构、参数，

进行动力学性能评估等；对改型的结构进行可制造、加工、装配分析；TLBI 需要其他智力单元的服务，不断利用 UDDI 注册中心和 AMT 提供的服务评估系统进行搜索和评估，寻找已经在 UDDI 中心注册的合作伙伴。

TEC、TLBI 和 AMT 初步形成合作联盟，并在设计进程中根据需要不断增加合作伙伴。

TLBI 提供异地合作设计管理平台（如图 49.6-5 所示），能够在权限控制下动态管理项目相关的各资源单元以及图形、数据的交换等。



图 49.6-5 基于 Internet 的异地合作设计管理平台

#### 4.4 主动电磁轴承的结构设计

TLBI 根据设计要求，进行电磁轴承支承概念的设计，通过调用本所开发的 Web 服务应用程序进行设计和计算，通过轴的结构设计，径向电磁轴承定子设计、径向电磁轴承转子设计、推力电磁轴承定子设计、推力电磁轴承转子设计，得到以 XML 格式保存的结构设计各参数；然后通过 VC 编写的 Solidworks 的 2D



图 49.6-6 电磁轴承结构设计的径向电磁轴承定子设计界面

次开发程序读取 XML 文件, 动态生成 SolidWorks 格式的三维 CAD 图形, 形成初步的设计结果。图 49-6-6 显示了电磁轴承结构设计的径向电磁轴承定子设计界面, 图 49-6-7 为以 XML 格式表示的电磁轴承结构设计结果, 图 49-6-8 显示了动态生成的 SolidWorks 格式的电磁轴承零部件三维图形。



图 49-6-7 以 XML 格式表示的电磁轴承结构设计结果

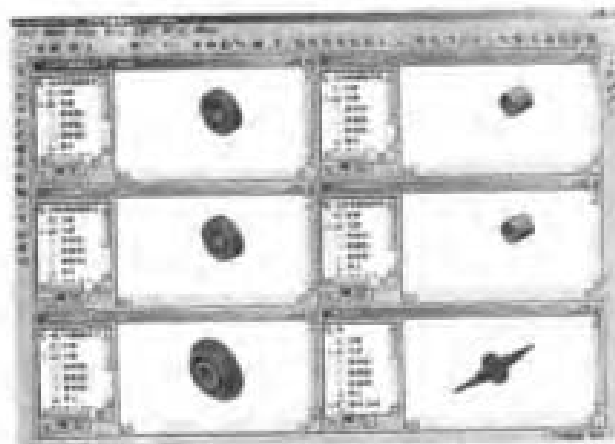


图 49-6-8 动态生成的 SolidWorks 格式的电磁轴承零部件三维图形

#### 4.5 转子轴承系统的动力学分析

由图 49-6-9 所示, 改型设计中, 去掉原有的润滑油系统和滑动轴承, 增加了传感器、控制器、功率放大器和执行器等部件。由于电磁轴承结构需要, 轴变长变细, 轴系的临界转速降低, TLBI 调用本所资源: Matlab 编写的转子轴承系统动力学分析程序进行动力学性能评估, 发现目前的设计方案无法满足临界转速要求, 经过再设计, 去掉原有设计方案中的辅助轴承, 减少轴的长度, 最终分析得出轴系的临界转速满足了设计要求。图 49-6-10 显示了基于 Internet 的转子轴承系统动力学 Matlab 分析程序界面和分析结果。

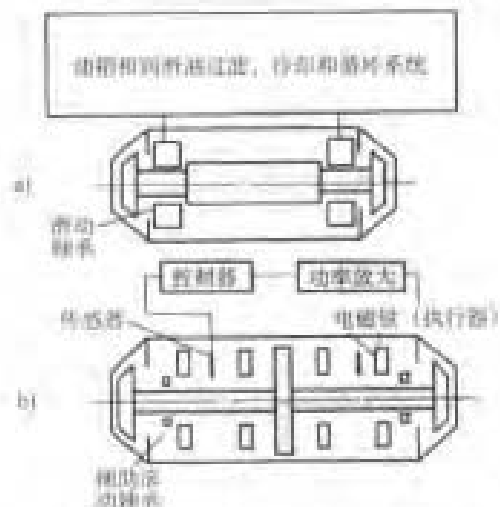


图 49-6-9 a) 采用动压滑动轴承支持的涡轮泵整机  
b) 采用电磁悬浮轴承的涡轮泵整机, 采用辅助滑动轴承进行掉电保护



图 49-6-10 基于 Internet 的转子轴承系统动力学 Matlab 分析程序界面

#### 4.6 涂层设计

由于原有的辅助轴承是起掉电保护的作用, 再设计提出了一个创新的概念, 即在电磁铁的表面涂上一层涂层, 以便在掉电时保护轴颈。查询 TLBI 的涂层网络数据库, 得到一种 TLBI 专用的涂层, 能够满足 40000r/min 和相应比压下的掉电保护的功能。图 49-6-11 显示了采用电磁悬浮轴承的涡轮泵整机应用辅助轴承和聚合物涂层来实现掉电保护功能的示意图。图中可以发现采用聚合物涂层后, 轴的长度缩短, 重新校验临界转速后得到目前结构可以满足临界转速要求。图 49-6-12 显示了涂层数据库查询界面。

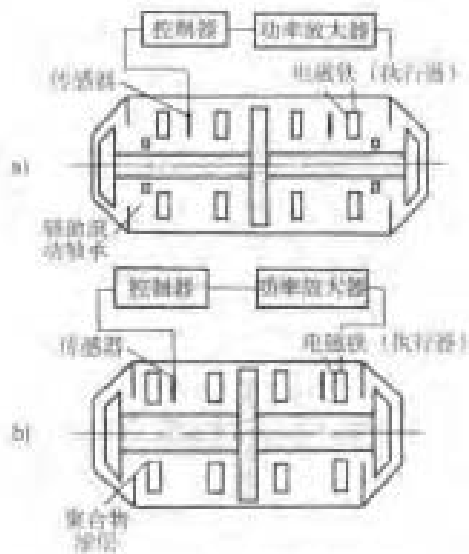


图 49.6-11 a) 采用电磁悬浮轴承的涡轮膨胀机, 采用辅助滚动轴承进行机电保护  
b) 采用电磁悬浮轴承的涡轮膨胀机, 采用在电磁铁表面涂上一层复合涂层



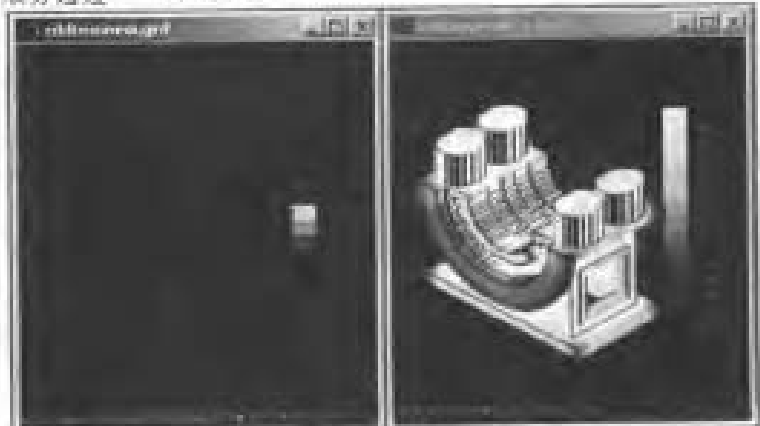
图 49.6-12 基于 Internet 的层次数据库查询界面

#### 4.7 与厂家交换设计意见

改型设计需要与请求方交换设计意见, 双方通过



a)



b)

图 49.6-14 a) 涡轮膨胀机底座的 UG 图形 b) 底座优化铸造工艺分析结果

Netmeeting 远程通信软件和 Communicator 远程图形交互软件和 TEC 在浏览器上对结果 (包括图纸) 反复进行网上对话, 如图 49.6-13 所示, 不同利益方取得了一致意见。



图 49.6-13 采用 Communicator 软件远程工程图形交互

#### 4.8 可铸造性评估

为满足资源单元知识服务对图形格式的不同要求, TLBI 又采用 UG 绘制了初步设计后的膨胀机的零部件图和装配图。初步设计的箱体十分复杂, TLBI 利用 UTDI 中心和服务评估系统搜索和评估能够提供铸造性分析服务的单位, 选择与清华大学机械工程系 (DME) 的服务合作。

TLBI 请求 DME 提供服务, 其通过异地合作设计管理平台下载相应的 CAD 文件, 图 49.6-14a 为涡轮膨胀机底座的 UG 三维图形, 进行涡轮膨胀机盖板、底座的可铸造性分析。

DME 进行涡轮膨胀机盖板、底座的可铸造性分析, 图 49.6-14b 为底座优化铸造工艺分析图。返回分析报告和设计改进意见提交到 TLBI 的异地合作设计管理平台。

#### 4.9 可加工性评估

TLBI 利用 UDFI 注册中心和服务评估系统进行搜索和评估提供可加工分析服务的单位, 选择华中科技大学工业工程与制造系统工程系 (DIMS) 合作。

TLBI 请求 DIMS 提供可加工性分析服务, 其通



图 49-6-15 a) 推力电磁轴承定子的 UG 图形, b) 推力电磁轴承定子机械加工工艺

过异地合作设计管理平台下载相应的 CAD 文件, 图 49-6-15a 为推力电磁轴承定子 UG 三维图形, 进行相关部件的可加工性分析。

DIMS 进行可加工性分析, 图 49-6-15b 为推力电磁轴承定子机械加工工艺, 返回分析报告和设计改进意见提交到 TLBI 的异地合作设计管理平台。

#### 4.10 可装配性评估

TLBI 利用 UDFI 注册中心和服务评估系统进行搜索和评估提供可装配分析服务的单位, 选择华中科技大学国家 CAD 支撑软件工程研究中心 (NERC-CAD) 合作; TLBI 请求 NERC-CAD 进行可装配性分析服务。

NERC-CAD 通过异地合作设计管理平台下载相应的 CAD 文件, 转换为 VRML2 格式的模型, 供网络环境下的协同装配分析, 实现了该产品基于 Web 的协同装配规划, 仿真和结果评估 (如图 49-6-16 所示), 最后返回虚拟装配分析报告和设计改进意见, 提交到 TLBI 的异地合作设计管理平台。



图 49-6-16 基于 Web 的协同装配规划, 仿真和结果评估

#### 4.11 制造

TLBI 汇总各方面的报告, 再设计的新知识, 对原来的概念设计进行修改, 最终形成可交付使用的设计, 对于各有关零部件, 在评估通过时, 热、冷加工的代码已经并行形成, 设计完成后, 立即就可以开始制造。

#### 4.12 基于 Internet 的远程试验

TLBI 开发的基于 Internet 的远程试验系统可以对模拟样机进行远程试验, 对设计进行最终评估。

#### 4.13 台架试验

图 49-6-17 为电磁轴承支承的涡轮增压机的整机装配图, 图 49-6-18 为电磁轴承—转子系统实物图, TEC 进一步的试验表明, 试验样机顺利实现升降速,



图 49-6-17 电磁轴承支承的涡轮增压机的整机装配图

运行稳定,性能符合要求。由此,我国已经具备生产高技术含量电磁轴承支承涡轮膨胀机的设计、制造生产能力,具有巨大的社会经济效益。

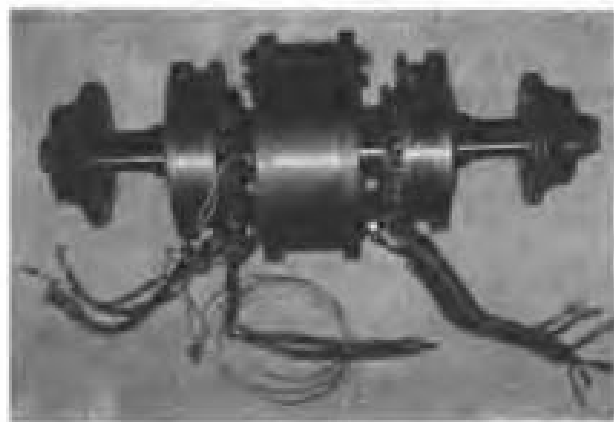


图 49.6-18 电磁轴承-转子系统实物图

### 参考文献

- Greenberg S. Collaborative Interfaces for the Web. In C. Forsythe, LEA Press, 241~254
- Grudin J. Computer-Supported Cooperative Work: History and Focus. IEEE Computer, 1994 (5): 19~26
- Haber G P. Issues in the design of group decision support systems. MIS Quarterly, 1984, 195~204.
- Bruck T, Gomez L. M. A Collaborative Medium for the Support of Conversational Proops. CSCW 1992, 171~178
- Monplaisir L. An integrated CSCW architecture for integrated product/process design and development. Robotics and Computer-Integrated Manufacturing, 1999; 15 (2): 145~193
- Lloyd P. Groupware in the 21st century: computer supported cooperative working now and the millennium. Greenwood, 1994
- Benjamin M. Groupware dynamics. Association Management November, 1995; 24
- Bidgen H. Group support systems: a new productivity tool for the 90's. Journal of Systems Management, 1995 (July/August); 36~62
- Korzaniowski P. Directory services slowly go global. PC Week, 1997 (June); 111~113
- Byrne J. Groupware fosters online collaboration. Government Computer News, 1997 (April); 29~30
- Lucero, Anthony, Yen, David C. a. Groupware, computer-supported collaboration. Telematics and Informatics, 1998, 15 (1~2); 85~101
- 袁清珂等. 产品数据表达与交换标准 STEP 的研究及应用. 机械科学与技术, 1997, 15 (6): 1097~1102
- 梁晓路, 梁宇奇. Web Services 技术: 架构和应用. 北京: 电子工业出版社, 2005
- Orfali R, Harkey D. Client/Server Programming with Java and CORBA. New York: John Wiley & Sons Inc., 1998
- 王怀民. 分布对象技术. 计算机世界, 1999, 15 (C): 1~2
- 魏长华. Agent 与面向 Agent 的程序设计. 华中师范大学学报(自然科学版), 1998 (3): 284~289
- 黄小兵, 唐文胜. 基于 Agent 系统的概念、方法和应用. 计算机与现代化, 2000 (4): 6~11
- Wooldridge N, Jennings N R. Intelligent agent: theory and practice. The Knowledge Engineering Review, 1995, 13 (2): 115~152
- Bondick D. Product Data Management, Enabling Enterprisewide Design Collaboration. CIM Strategic Analysis Report. Gartner Group, 1995
- CIMdata, Inc. Product Data Management. A Technology Guide, 1996
- 谢友柏. 现代设计与知识获取. 中国机械工程, 1996, 7 (6): 36~41
- 胡亚红, 周受欽, 姜培林等. 现代设计中的远程程序调用 [J]. 中国机械工程, 2001, 12 (9): 1035~1037
- 周受欽, 薛安, 谢友柏. 滚动轴承知识表征与检索方法 [J]. 西安交通大学学报, 2001, 35 (2): 175~178
- 康荣学, 贾海波, 崔优云. 基于 Internet 的远程实验研究 [J]. 计算机工程与应用, 2002 (16): 168~170
- 刘刚. 异地合作设计服务研究与实现 [D]. 西安: 西安交通大学机械工程学院, 2002
- 谢振宇. 电磁轴承系统及其工业应用 [D]. 西安: 西安交通大学机械工程学院, 2000

# 附录 现代设计主流软件

主 编 蒋祖华  
编写人 蒋祖华



# 第1章 CAD/CAM 软件

## 1 UGII 软件

### 1.1 UGII 软件概述

UGII 软件是美国 Unigraphics Solutions 公司的 CAD/CAM 产品,其核心 Parasolid 提供强大的实体建模功能和无缝数据转换能力。UG 实现了全相关的和数字化实体模型之间的数据共享,它提供用户一个灵活的复合建模模块,如实体建模、曲面建模、线框建模、基于特征的参数建模以及功能强大的逼真照相的渲染、动画和快速原型工具。UG 着眼于部件级的 GTX-T,它使用户能快速和精确地通过公差特征将公差信息与几何对象相关联。强大的编程框架使用户和软件供应商可以开发出与 UG 能很好集成并全相关的应用程序。UG 覆盖制造全过程,以及制造的自动化、集成化和用户化,在产品制造周期、产品制造成本和产品质量方面,都提供了实用的、柔性的 CAM 产品,融合了工业界丰富的产品加工经验。图附-1-1 为可变轴视图。



图附-1-1 UG 可变轴视图

### 1.2 UG 软件功能介绍

#### 1.2.1 产品设计

##### (1)UG/入口(UG/Gateway)

UG/Gateway 是连接所有 UG 模块的基础。它支持关键操作,包括打开已存在的 UG 部件文件、创建新的部件文件、绘制工程图和屏幕布局以及读入输出各种不同格式的文件,包括参数实体(Parasolid)转换文

件和计算机图形的图档文件(CGM)。同时提供的还有层控制、视图定义、屏幕布局和显示控制功能,包括消除/再现对象和在线帮助功能。UG/Gateway 还提供一个与各种高分辨率绘图机接口的许可证,并提供先进的电子表格应用。通过 UG/Gateway 的导航、动画、实体和曲面模型的显隐与着色,提供了高级的可视化功能。建造和管理零件族,操纵零件间的表达式,以便通过分析相关联的方案方便地扩充模型设计。空间漫游,可以定义漫游路径,生成电影文件。表达式查询;特征查询;模型信息查询;坐标查询;距离测量;曲线曲率分析;曲面光照分析;实体物理特性自动计算;用于定义标准化零件族的电子表格功能;按可用于互联网主页的图片文件格式生成 UG 零件或装配模型的图片文件,这些格式包括:CGM、VRML、TIFF、MPEG、GIF 和 JPEG;输入、输出 CGM、UG/Parasolid 等几何数据;Macro 宏命令自动记录、回放功能;User Tools 用户自定义菜单功能,使用户可以快速访问其常用功能或二次开发的功能。标准表格查询功能提供了一种简单的方法来实现基于知识的工程技术。

##### (2)UG/实体建模(UG/Solid Modeling)

该模块将基于约束的特征建模和显式几何模型方法无缝地结合起来,提供了当今业界最强有力的“复合建模”工具,使用户可以充分利用集成于先进的、基于特征环境中的传统的实体、面、线框造型的优势。UG/Solid Modeling 可使用户能够很方便地建立二维和三维线框模型、扫描和旋转实体以及进行布尔运算及参数化编辑。它还提供用于快速、有效的概念设计的变量草图工具和更通用的建模和编辑任务的工具。该模块易于理解的基于图标(icon)的图形环境与其他建模模块的操作相同。

##### (3)UG/特征建模(UG/Features Modeling)

该模块提高了表达式设计的层次,设计信息可以用工程特征的术语来定义。它提供了支持建立和编辑下列各种标准的设计特征:孔、槽、型腔、凸台、垫、柱体、块体、锥体、球体、管状体、杆、倒圆和倒角等。同时还可以挖空实体建立薄壁件。特征还可以被参数化定义并对其尺寸大小和位置做尺寸驱动编辑。用户自定义的特征存储在公共目录下,也可以被添加到设计模型中。特征还能够相对于任一其他的特征或物体定位,也能够被引用以建立相关特征组。特征组排列可以是个别的定位,也可以是在简单图案和阵列中定位。

#### (4)UG/自由曲面建模(UG/Freeform Modeling)

该模块支持复杂的自由曲面的形状,如机翼、进气道和其他工业产品的造型设计。它可将实体建模和表面建模的技术合并,建成一个功能强大的建模工具组。此建模技术包括沿曲线的通用扫描法,使用一条、二条和三条轨道方法按比例地建立外形,用标准二次曲线的方法建立二次曲面体,建立圆形或二次以及圆锥截面的倒圆(圆角)面,光滑桥接在二个或更多的其他实体间间隙的曲面。对于逆向工程,UG/Freeform Modeling 也可通过曲线/网格来定义形状或者通过点云来拟合形状,模型还可以通过修改所定义的曲线、改变参数值或使用图形和数学规则来控制进行修改编辑,如一个变半径倒角或变截面扫描功能。由此而建立的模型与所有其他 UG 功能完全集成为一体,UG/Freeform Modeling 还提供计算复杂模型的形状、尺寸和曲率的易于使用的工具。

#### (5)UG/用户自定义特征(UG/User-Defined Features)

该模块提供一种利用用户自定义的特征(UDF)概念去捕捉和存储零件族的方法,易于恢复和编辑。该模块还提供了一些工具,如允许取一个已存在的用标准 UG 模型工具建立的参数化实体模型,建立特征参数之间的关系、定义特征变量、设置默认值以及确定调用特征时所采用的一般形式所需的全部工具。用户自定义特征建立后,将驻留在一个目录中,在此目录上,供使用该模块的任何用户访问。当 UDF 已加入到设计模型中后,使用常规的特征编辑方法就可以对模型的任何参数进行编辑,使其性能符合原始 UDF 生成器建立的设计意图。

#### (6)UG/工程制图(UG/Drafting)

UG 工程制图模块提供了自动视图布置、剖视图、各向视图、局部放大图、局部剖视图、自动、手工尺寸标注、形位公差、粗糙度符合标注、支持 GB、标准汉字输入、视图手工编辑、装配图剖视、爆炸图、明细表自动生成等工具。该模块可以使设计者、工程师和绘图员从三维实体模型得到完全相关的二维工程图。利用 UG 的复合建模技术,该模块可生成尺寸与实体模型相关的工程图,并保证随着实体模型的改变而同步更新工程图尺寸,减少了用模型二维图样更新所需的时间。包括消隐线和相关横截面视图在内的二维视图在模型修改时也会自动更新。自动视图布局功能可快速布局二维视图,包括正交视图、轴侧视图、剖面图、辅助视图和局部放大视图等。通过重新利用模型数据,在图上可以自动地沿用相关的模型尺寸和 GD&T 符号,大大节省了时间。UG/Drafting 支持工业上发布的主要制图标准,如 ANSI/ASME、ISO、DIN 和 JIS,并提供了一套完整

的基于图标菜单的绘图和标注工具。利用 UG/Assembly Modeling 的数据可以方便地绘制装配图,并能快速生成装配分解图。无论绘制单页还是多页详细装配图及零件图,UG/Drafting 都能减少绘图时间和成本。

#### (7)UG/装配建模(UG/Assembly Modeling)

该模块提供了并行的、自上而下的产品开发方法。UG/Assembly Modeling 的主模型在整个装配过程中可以进行设计和编辑。部件可灵活地配对或定位,并且一直保持其关联性。这样既改进了性能,又节约了磁盘的存储空间。装配件的参数化建模还可以描述各部件之间的配对关系、确定通用紧固件组及其他复制的部件。这种体系机构允许建立非常庞大的产品结构并为设计组之间共享,使产品开发组成员始终与他人并行地工作。按用户规定的命名规则或采用 UG/Manager 的配置规则,都能够正确地访问不同版本的零件。

#### (8)UG/高级装配(UG/Advanced Assemblies)

该模块将高速渲染着色和间隙分析技术相结合,提供了数据装载控制功能,允许用户对装配结构的部件进行过滤分析。UG/Advanced Assemblies 可以管理、共享和评估数字模型,以完成一个复杂产品的全数字化装配模型。用它提供的各种工具,用户可对整个产品、指定的子系统或本件进行可视化及装配分析,并使产品的性能和生产率达到最优化。变更模型的表现方式,允许迅速地进行间隙检测,并对有阴影和隐藏线的视图着色。间隙的检测结果可被保存以备将来使用,并且有选择地在批处理方式下运行。如果需要,该模块对硬干涉也可给出精确的答案。当一个大型产品的部分结构作工程改变时,可以定义其区域和组件集,并由设计团队共享,以提高响应速度。

#### (9)UG/虚拟现实(UG/Reality)、UG/漫游(UG/Fly-Through)

这些模块提供了分布式工具、并行可视化工具和虚拟产品模拟化的工具。这些模块利用 UG/Advanced Assemblies 来精确显示和进行动态干涉检验。UG/Reality 对产品是否适用及功能性方面进行实时模拟的同时对产品进行评估。此模块利用 UG/Reality 技术根据部件的运动、装配的步骤和部件内部的漫游与 UG/Fly-Through 一起来建立动画。UG/Reality 允许建立运动副、显示联接处的滑动或转动部件,并且模仿真实运动,对装配行为和装配顺序提出建议,UG/Fly-Through 技术可利用 UG/Reality 提供迅速的易于使用的导引和虚拟产品的可视化进行部件的运动过程动画重放。多个用户可以同时观察虚拟产品并可与其交互,还可以和其他部门一起设计进行协作评审。

#### (10)UG/逼真着色(UG/Photo)

该模块提供高级图形工具,包括可选择的质量等级、视图着色、装配图着色、动画、正交视图和透视图着色、光源、阴影和工程材料库,从而加强了 CAD 模型的视觉效果。它是一个非常直观的工具,用于帮助在设计人员之间、设计者与客户或供应商之间进行设计意图和设计思想的沟通,如提供设计评审、产品建议、用户意见和市场资料等。UG/Photo 模块与 Unigraphics 的其他模块是完全集成的。

#### (11)UG/标准件库系统(UG/FAST)

该模块提供了一个用 UG 的三维实体格式定义的标准件库系统,并且很容易通过直观的图形界面来获得这些标准件。此库系统提供广泛的三维标准件库:ISO 紧固件、ANSI 紧固件、DIN 紧固件、DIN 轴承、DIN 钢结构和一个 UG/Fast 用户化工具包,以用来处理特殊用户的信息。UG/Fast 支持 UG/Assemblies、UG/User-Defined Features 或零件的读入。这个模块可以使标准件具有子装配功能,并且可被封装到 IMAN 和 UG/Manager 中。

#### (12)UG/WAVE 技术

UG/WAVE 技术提供了一个参数化产品设计的平台,为了维持设计的完整性和意图,此技术把概念设计与详细设计的变化自始至终地贯穿到整个产品的设计过程。在此平台上具有创新的 WAVE 工程技术,使其高级产品设计的定义、控制和评估成为可能。这一被称为“控制结构”的可重复利用的设计模板被用来表达产品设计的概念,这是通过定义几何形体框架和关键设计变量来实现的,CAE 的需求也被考虑到模型中,PDM 则被用来管理这些变动和版本。通过参数化的编辑控制结构,不同的设计概念可以被迅速地分析和评估。在控制结构中的关键几何模型可被相关联地拷贝到经过详细设计的产品装配中,这样在今后的产品开发过程中就允许高级概念设计的变化与整个产品设计的改变相关联。

#### (13)UG/几何公差(UG/Geometric Tolerancing)

该模块由于具有模型全相关件,故使智能化定义几何公差成为可能,它提供了尺寸公差和变量分析的基础。基于选择的公差标准,如 ANSI Y14.5M-1982、ASME Y14.5M-1994 和 ISO 1101-1983。通过一个很方便的用户界面,该模块可支持简单而迅速地对基准和公差建立编辑、数据查询、基准和公差的句法和相关检验的过程,并且根据模型或数据的改变自动更新公差,并自动在图样上沿用 GD&T 符号。通过一个综合的 UG/OPEN API,各部件的公差分析、装配件公差综合分析和检查之间的信息可传往下游应用进行,在模型中装入公差的益处在于协调和编译的 GD&T,消除多余的输入数据,减少对图样的依赖性,获得正式的

文件,重新审查产品需求以及方便的工程制造分析。

### 1.2.2 产品制造

#### (1)UG/CAM 基础(UG/CAM Base)

该模块提供了连接 UG 所有加工模块的基础。用户可以在图形方式下通过观察刀具运动,用图形编辑刀具的运动轨迹,具有延伸、缩短和修改刀具轨迹等编辑功能。针对如钻孔、攻螺纹和镗孔等加工任务,它还提供了通用的点位加工程序,用户化对话特征允许用户修改对话和建立适于它们的专用菜单,这就减少了培训时间和流水线加工作业工步。通过使用操作模板可进一步提高用户化水平,如允许用户建入粗加工、半精加工等专门的样板子程序,常用的加工方法和工艺参数都已标准化。

#### (2)UG/后置处理(UG/Postprocessing)

该模块使用户对工业上的大多数 NC 机床很容易地建立自己的后置处理程序。UG/Postprocessing 的功能包含了铣加加工(2~5 轴或更高)、车加工(2~4 轴)和线切割加工等实际应用的检验。

#### (3)UG/车加工(UG/Lathe)

该模块提供了高质量返转类零件加工所需的全部功能。零件的几何模型和刀具轨迹完全相关,刀具轨迹能随几何模型的改变而自动更新。它具有粗车、多次走刀精车、车沟槽、车螺纹和中心钻孔等功能。输出的刀位源文件可直接进行后处理,产生机床可读的输出文件。用户可控制的参数有进给速率、主轴转速和零件间隙等。若不作更改,这些参数将保持原有数值。通过屏幕显示刀具轨迹,对数控程序进行模拟,便可检测设置参数是否正确。文本输出生成一个刀位源文件(CLSF),对刀位文件用户可以存储、删除或按要求修改到正确位置。

#### (4)UG/型芯和型腔铣削(UG/Core&Cavity Milling)

该模块对汽车和消费品行业中加工模具和冷冲模特别有用。它提供粗切单个或多个型腔、沿任意形状切去大量的毛坯材料以及可加工出型芯的全部功能。最突出的功能是对非常复杂的形状产生刀具运动轨迹,确定走刀方式。容差型腔铣削可用于加工不精确的设计形状的曲面之间有间隙和重叠的场合。可被分析的型腔面数目多达几百个。当 UG/Core&Cavity Milling 检测到异常的型腔面时,或是对它修改,或是在用户规定的公差范围内加工出型腔。该模块提供了型芯和型腔加工过程的全自动化。

#### (5)UG/固定轴铣削(UG/Fixed Axis Milling)

该模块提供了完全和综合的工具,用于产生 3 轴运动的刀具路径。实际上,它能加工任何曲面模型和实

体模型。可以用功能很强的方法来选择零件需要加工的表面或加工部位。它有多种驱动方法和走刀方式可供选择,如沿边界、径向、螺旋线以及沿用户定义的方向驱动,在边界驱动方法中又可选择同心圆和径向等多种走刀方式。此外,它还控制逆铣和顺铣切削以及沿螺旋路线进刀等。同时,还可容易地识别前道工序未能切除的区域和陡峭区,以使用户进一步清理这些地方。UG/Fixed-Axis Milling 可以仿真刀具路径,产生刀位文件,用户可接受并存储此刀位文件,也可拒绝或按要求修改某些参数。

#### (6)UG/清根切削(UG/Flow Cut)

UG/Flow Cut 处理器能节省半精加工或精加工的处理时间。这一模块同 UG/Fixed-Axis Milling 功能模块结合起来,分析零件的加工面(以加工参数为基础),同时检测所有双相切条件。这些区域一般存在于型腔的凹谷处和拐角处。用户可以指定刀具,利用双相切(条件)来定义驱动轨迹。该处理器模块将自动在这些区域用一次走刀或多次走刀移去未被切除的材料。当加工复杂的型芯和型腔时,此模块在精加工刀轨作用之前减少了精加工零件表面或获得均匀余量。

#### (7)UG/可变轴铣削(UG/Variable-Axis Milling)

该模块提供任何 UG 曲面的固定轴和多轴铣削的功能。规定了 3~5 轴轮廓运动、刀具定向和曲面加工质量。通过用曲面参数把刀具轨迹映射到加工面上,并利用任意曲线及点就可以对刀具轨迹进行控制。

#### (8)UG/顺序铣切削(UG/Sequential Milling)

该模块适用于用户要求对切削过程中刀具的每一步路径生成都要进行控制的情况。UG/Sequential Milling 和几何模型是全相关的。它类似于从前市场上 APT 系统处理的软件,但生产效率要高得多。用交互式可以逐段地建立刀具路径,但处理过程的每一步都受总控制的约束。一个称为循环(Looping)的功能允许用户通过定义轮廓的里边和外边轨迹后,在曲面上生成多次走刀加工,并可生成中间各步的加工程序。

#### (9)UG/制造资源管理系统(UG/Genius)

该模块能方便、高效地建立制造数据并加以分类。功能强大的关系数据库系统特别适合于支持生产计划、刀具、NC 程序、订货数据和库存管理等功能。UG/Genius 基于模块化原理、易于扩充以适应各种各样用户的需要。它还具有如下许多特点,诸如可提供图形刀具分类、具有与 UG/CAM 的接口以及各种 MRP、DNC 系统的接口。

#### (10)UG 切削仿真(VERICUT)

该模块是 CGTech 公司开发的、内嵌于 UG 的软件。它采用人机交互方式可模拟、检验和显示 NC 刀具的路径,是一种花费少、效率高的不用机床就能验证数

控程序的好方法。由于省去了试切样件,可节省机床调试时间、减少刀具磨损和机床清理工作。通过定义被切零件的毛坯形状,调用 NC 刀具轨迹数据,就可以检验由 UG 生成的刀具路径的正确性。Vericut 可以在工作台上显示出加工后并着色的零件模型,以使用户观察是否出现了不正确的加工情况。作为检验的一部分,该模块还能计算出完工零件的体积和毛坯的切除量,因此很容易确定原材料的损失。Vericut 还提供了许多其他功能,包括对毛坯尺寸、位置和方位的完全图形控制以及模拟 2~5 轴联动的铣削和钻削加工。

#### (11)UG/线切割(UG/Wire EDM)

该模块支持 UG 的线框模型和实体模型,可进行 2~4 轴线切割加工。在编辑和更新模型时,所有的操作是全相关的,它提供了多种线切割加工方式。如多次走刀的轮廓加工、电极丝反转和切割留有成块材料的加工。同时它也支持定程切割以及使用不同直径的电极丝和功率大小的设置。用户还可以用通用的后处理器来开发专用的后处理程序,生成适用于某个机床的数据文件,UG/Wire EDM 也支持流行的 EDM 软件包,如 AGIE、Charmilles 及其他软件。

#### (12)UG/图形刀轨编辑器(UG/Graphical Tool Path Editor)

这个 CAM 模块可以让用户观察到刀具沿其轨迹运动的情况,用户能够通过操纵图形和文本信息来编辑刀具轨迹。编辑器然后显示出一个编辑修改后生成的刀具轨迹结果。该模块还提供了刀具动画功能工具,可以在整个或部分刀轨段上显示动画,同时还可控制动画的速度和方向。另一个重要特点是对已经被限定了边界的刀具轨迹进行延伸和裁剪,如压板或夹具所限定的边界,并且能进行过切检查。

#### (13)UG/机床仿真(UG/Unisim)

该模块为用户提供了一个功能强大的可视化系统,此系统是为提供一个“逼近现实”的加工仿真环境设计的。其目的是为了在复杂的加工环境中减少加工时间、消除机床损坏并提高质量。UG/Unisim 包容了整个加工环境——机床、刀具、夹具和工件,以用来仿真,同时也是为了检验的目的。通过使用从 UG/CAM 中得到的后处理的输出数据,UG/Unisim 可以精确地检测相互接触的部件之间的碰撞。

#### (14)UG/SHOPS

这是一个针对车间的可升级的加工产品家族。UG/SHOPS Die/Mold 是一初始产品,主要用于车间生产中的冲模和压模的加工。用户可以通过一个基于工作台对话的、易于使用的图标来一步步地移动导航,一旦几何模型确定下来,它就可以在随后的加工命令中被自动使用。用户可以选择毛坯的几何模型来进行

原始粗工。工作台包含已被证实的粗加工和细加工的加工方法,用户可以根据刀具材料、零件材料、刀具尺寸和外形来选择进刀速率和速度。对于型芯、型腔和清根切削加工,UG/SHOPS 提供了一整套粗加工处理方法。曲面加工处理器允许用户进行高质量的精加工。一个可视化模块也被并入此产品中,这样用户就可以显示已生成的刀具轨迹的结果。

(15) Nurbs (B 样条) 轨迹生成器 (Nurbs (B-Spline) Path Generator)

该模块允许从 UC/NC 处理器中直接生成基于 Nurbs 的刀具轨迹数据。直接从 UG 实体模型中获得的新的刀具轨迹可以加工出极其精确和超等级的零件。通过消除控制器等待时间,用户可看到物理纸带尺寸被减小到标准格式的 30%~50%,加工时间也明显地减少。如果用户想充分利用新的高速机床的优点,而这些高速机床又提供功能强大的控制器特征的话,那么 UG/Nurbs (B 样条) 轨迹生成器就是一个用户所必需的工具。

### 1.2.3 产品分析

(1) UG/有限元分析 (UG/Scenario for FEA)

该模块是一个集成的、相关的、直观的 CAE 工具,它能快捷地对 UG 的零件和装配部件进行前、后处理。该产品作为设计过程的一个集成部分,用于评估各种设计更改方案或“Scenario”的性能。这种分析结果是一个工程预测的过程,它可以优化产品设计、提高产品质量、缩短产品上市的时间。用于有限元分析的 UG/Scenario for FEA 提供了将几何模型转换为有限元模型以及图形化评估分析结果的能力。一个可选的集成的求解器 UG/FEA,可以进行线性静力分析、模态分析和稳态热分析。

(2) UG/有限元 (UG/FEA)

该模块是一个有限元分析求解器,它与 UG/Scenario for FEA 的前、后处理能力紧密地集成在一起。这两个产品为在 Unigraphics 环境里进行建模与分析提供了一个完整的解决方案。UG/FEA 为概念设计提供了广泛的求解类型,包括线性静力、标准模态、稳态热传递和线性屈曲分析,同时还支持装配部件,包括间隙单元的分析。对薄壁结构和梁的尺寸优化是一种可选的功能。UG/FEA 支持的材料类型有正交各向同性、正交各向异性 and 各向异性。另外,在求解中可以定义与温度相关的材料特性。

(3) UG/机构学 (UG/Mechanisms)

该模块能很方便地在 UG 中对任何实际的二维或三维机构进行复杂的运动学分析和设计仿真,可以完成大量的装配分析工作,诸如最小距离、干涉检查、轨

迹包络等。一种独特的交互式运动学模式允许同时最多控制五个运动副。用户可以分析反作用力并图解合成位移、速度、加速度。反作用力可输入到 FEA 模块中。UG/Mechanisms 支持丰富的机构运动单元库。几何模型可以用来定位运动副、力及定义 CAM 轮廓。该模块嵌入的是 Mechanical Dynamics 公司 (MDI) 的求解器 ADAMS/Kinematics。同时对于复杂问题,它能为 MDI 的全部动态求解器 ADAMS/Solver 建立输入文件。

(4) UG/注塑模分析 (UG/ MF Part Adviser)

该模块是一个易用的、全塑流的注塑模分析系统,它集成在 UG 中,具有前处理、解算和后处理能力,并提供了在线求解器和完整的材料数据库。分析结果将动态显示注塑过程中的流动、填充时间、焊线位置、气井、填充的可靠度、注塑模压力和降温过程。使用该模块可以帮助模具设计人员确定其注塑模设计是否合理,不合适的注塑模几何体会被很容易地检查出来并予以修正,从而生产出高质量的注塑模。

### 1.2.4 钣金件 (Sheet Metal)

(1) UG/钣金制造 (UG/Sheet Metal Design)

该模块包含一组成形设计特征,它是专门为带直线弯边的钣金产品的展开、压模、剪切定制的。这些特征使设计人员能够以准确的变形图来定义和模拟加工工序——未成形、部分成形和再成形模型。建立在其他成形特征上面的钣金特征将按基本特征,弯边/不去边来变形,用户定义的形状可用来建立成形的剪切口与模压。在建立新的特征时,可应用并自动检查用户标准。

(2) UG/钣金制造 (UG/Sheet Metal Fabrication)

对用 UG/Modeling 软件设计的钣金件,此模块提供了从转塔式多工位冲压到激光切割的功能。也可对带圆孔和矩形孔特征的钣金件冲压进行自动编程或者用户对冲压操作进行交互式编程。用户还可以在一张金属板上以栅格形式排列加工程序的拷贝。

(3) UG/钣金件排样 (UG/Sheet Metal Nesting)

该模块可在一块毛坯板材上对若干品种的部件进行多种优化排样。用户对参加组合排样的零件种类、每种零件的数量以及所使用板材的规格进行选择。“自动排样”功能允许用户对控制全部排样件的布置进行不同的组合布置策略的择优选择。该模块还能优化刀具更换和冲压工序,使冲压零件时的板材重定位次数最少。用户还可用交互式图形方式直接在板材上进行排样。Unigraphics Solution 的专家还可以为用户开发专用的钣金加工后处理程序。

(4) UG/高级钣金设计 (UG/Advanced Sheet

## Metal Design)

当对一个复杂的钣金产品进行压模、拉模和成型操作时,UG/Advanced sheet Metal Design 提供了一套定制的成型设计特征和工具。这些特征,典型的如曲线弯边边缘,可在汽车、航空航天及消费产品中见到。“桥”特征即可以建在参考面、参考边缘和参考曲线上;也可以从参考面、边缘、曲线中来建立。这些参考几何体允许高级钣金特征建立在部件间的几何体上,或是从原先没有定义钣金特征的模型中来建立。UG/Advanced sheet Metal Design 模块使用各种与材料类型和成型操作相匹配的求解器,去建立复杂曲面的展平表示,提供了一个展平图样工具。展平图样工具不仅限于在钣金设计特征中使用。

## (5) UG/钣金冲模工程(UG/Sheet Metal Die Engineering)

该模块为冲压操作中设计冲模面提供了一组建立可成形裁剪边缘、边料、组合件的工具。该工具对于成形过程中分析模具截面、边料和组合件也是有效的。软件也提供了用于建立准确的模具边缘和用于确保清边而检查裁剪角的工具。

## 1.2.5 用户化——UG/Open

## (1)UG/Open MenuScript

该模块为用户提供了利用定制菜单条和弹出菜单来用户化 UG 的能力。用“add on”菜单文件可完成广泛的定制层,包括重排、移去菜单文件及向 UG 中增加用户功能。定制的调用函数、UG/Open 应用程序、系统命令和工具面板可逻辑地成组在菜单中。利用该模块用户可以将第三方产品和企业应用程序非平行地集成在 Unigraphics 中。

## (2)UG/用户界面设计(UG/Open UIStyler)

该模块使构造用户界面的传统手工方法自动化,这样使用户就能快捷地构造 UG 风格的对话框。用户可在简单、易用的图形编辑器里从大量的标准 UG 用户界面对象选择中来构造对话框,通过这种能力来生成模板代码、定义对话框行为、集成为特殊应用编写的程序和利用已有的 Unigraphics 工具,大大提高用户的效率。该模块和对话框完全独立于硬件平台,在 Windows NT 中自动采用类似于 Windows 的窗口,在 Unix 中自动采用类似 Motif 的窗口。

## (3)UG/Open GRIP

该模块是实现 CAD/CAM/CAE 任务自动化的中间描述语言,用户可以编写程序实现自动数控(NC)操作,也可以建立几何体和制图对象,自动建模和制图。

## (4)UG/Open API

该模块提供了与 UG 的直接编程接口,允许用户

使用当今流行的大多数编程语言,如 C、C++、Java 建立自己的应用程序。UG/Open 应用程序接口(API)通过使用自定义对象,提供了一个“全扩展数据模型”。这就使 UG/Open 应用程序能定义新类型的 UG 对象,它可被显示、相关连接到标准的 UG 对象,并永久存储在 UG 的部件文件中,可以开发应用程序扩展 UG,也可以充分利用 UG 内的构造功能建立完全独立的实用工具。

## (5)UG/Open++

UG/Open++ 是一个真正面向对象的与 UG 的接口。这个用 C++ 编写的 API 充分利用了诸如继承、封装和多形态这些面向对象的特征。UG/Open++ 提供了完全存取其分类层次,允许用户重构、引用自己的分类,在 UG 中建立新的、永久的对象。UG/Open++ 与 UG/Open API 是完全兼容的。为复杂编程任务提供了一个简单、易用的解决方案。

## 1.2.6 Web 产品

## (1)UG/web 服务器(UG/Web Server)

该模块为用户提供了—套工具,该工具由一个 Web 浏览器的客户端提供控制,将 UG 连接到 Web 服务器上。UG/Web Server 利用标准的 UG/Open 开发工具和 Web 构架可支持非常广泛的应用,如通过浏览器来修改 UG 模块等。

## (2)UG/Web 表示(UG/Web Express)

该模块允许用户利用 UG 零件和装配件数据来创建 Web 页面。它提供了一组丰富的内嵌命令集,使用户可抽取 UG 的数据,如属性、表达式和工程电子表格信息等数据。它内含预定义的 UG HTML 模板,从而允许快捷地创建用户化模板。通过 UG/Open API 可以存取 UG 的数据并加到一个 Web 页面上。UG/Web Express 提供了一种工具,使用 Internet 标准与 CAD/CAM/CAE 的通讯技术以使用户很容易地将 UG 的模型数据发布到 Web 页面上去,并可从任何 Web 浏览器上查看生成的资料。

## (3)UG/Web 渲染(UG/Web Render)

该模块允许用户以各种不同的 Web 支持的格式生成 UG 模型的图像文件;如 CGM、VRML、TIFF、MPEG、GIF 及 JPEG。VRML Export 支持在输出文件中加插图。UG/Web Render 是 UG/Gateway 的一部分。

## (4)UG/GenConnect 基于知识的工程(Knowledge Based Engineering)

该模块提供了支持交互式存取 ICAD 系统的接口。用户可以采用 UG 几何模型选项或以菜单选项的形式输入建模参数。ICAD 系统中定义的规则将根据

这些设计参数自动生成产品结构,并传送到 UG 部件中去。利用 ICAD 设计语言(IDL),用户可以将设计经验建成各种规则,用这些规则来定义一个完整的产品数据结构,所有参数如成本、重量、加工性能或生产率都可以进行优化,当产品的技术规格改变时,规则将自动作出必要的改动。

### 1.2.7 管路应用(Routing Applications)

(1) UG/Routing, UG/Tubing, UG/Piping, UG/Conduit, UG/Raceway, UG/Steelwork。

该模块为所有 UG 管路应用提供厂共同的用户接口和用户工具。它加快了对走线应用,如管道、管路、导槽、导线、电缆管道、水道和钢结构装配件的建立,同时为下一步应用捕捉所需信息。最终的智能装配件可以进行间隙分析、生成生产图样、产生材料清单、驱动用户分析和制造应用。UG/Routing 可以和用户自己的标准部件、设计规则和系统接口一起使用,或是与特定的走线模块,如 UG/Tubing、UG/Piping、UG/Conduit、UG/Raceway 及 UG/Steelwork1 一起使用。

(2) UG/电气配线(UG/Harness)

该模块可在复杂的装配件内自动完成电气配件的设计和加工过程。UG/Harness 从原理图捕捉软件包,如 LTX 软件公司的 UG/Schematics 中读取网线表,并查找 UG 装配件中的设备连接关系。然后 UG/Harness 能够精确计算三维导线长度、估算电气配线的线束直径。UG/Harness 将生成线束的三维表示,以进行间隙分析。UG/Harness 将展平三维电气配线,用于为制造产生与实际尺寸一样的夹具模板或安装布局。

### 1.2.8 质量工程应用(Quality Applications)

UG/设计检验、UG/检测、UG/逆向工程(UG/Design Verification, UG/Inspection, UG/Reverse Engineering)这些质量工程应用模块可使工程师对设计意图进行精确定义、解释和交流,以使设计的产品达到优质、高效。实行闭环质量控制可以节省 70% 测量机的编程时间和 80% 的返工时间,同时还可以节省 40% 的加工调整和装夹时间。这些内嵌的软件工具可以通过 Unigraphics Solutions 公司与 Tecnomatix 质量工程公司(Tecnomatix Technologies 公司的子公司)和 VSA 系统分析公司得到。

### 1.2.9 数据交换(Data Exchange)

UG 数据交换产品提供了强大的模型和制图信息的交换功能。所有读入到 UG 中的数据都是可直接利用的。UG 基于标准的 IGES 和 STEP 产品被公认为在

数据交换方面处于世界领先地位。UG 还提供了大量的直接转换器(CATIA、CADD5、SDRC、EMC 和 AUTOCAD)以确保同其他系统高效地进行数据交换,最大限度地提供了将数据转换到较弱的二维系统中的能力。UG 数据交换定位在访问数据交换中最重要方面,即保留传统数据、消除重复工作费用、具有同所有输入数据完全的相互可操作性。

### 1.2.10 特殊应用(Special Applications)

(1) UG/快速成型(UG/Rapid Prototyping)

该模块为生成快速成型机信息提供了高效、易用的能力。它能生成工业标准的“STL”格式文件,该文件允许用户在一个加速的从艺术品到产品的环境中,快速评定设计。该模块允许用户以精确实体或曲面模型为工作对象,它将所有的需要转换为 STL 表示,也能使用户以要用的快速成型机作为基础控制公差。

(2) UG/管理者(UG/Manager)

UG/Manager 增加了对 Unigraphics 应用多用户的协调和产品数据管理功能,以便使工作组的成员能够较好地协作。它与 UG/Assembly 建模模块完全集成在一起。

(3) UG/移植(UG/MX)——CADAM 移植(CADAM Migration)

该模块为运行在大型机上的 CADAM 软件提供了将二维设计/工程制图转换为带有实体模型的三维应用的低成本、高效率的移植手段。UG/MX(Unigraphics Migration Extension)能把运行在大型机上的 CADAM 软件移植到网络工作站上或与 CADAM 进行图形交换。在同一台工作站上支持由 CADAM 软件形成的传统二维工程图和 UG 的二维实体设计、制图、分析、制造等功能,该模块具有完备的移植辅助功能。

## 2 Pro/Engineering 软件

### 2.1 概述

Pro/Engineering 软件是美国 PTC 公司的产品,是一种单一数据库结构、标准的基于特征的建模技术以及独特的全相关技术的机械设计自动化软件。其特点有:

1) 先进的特征建模技术:Pro/ENGINEER 2000I 扩展的特征建模工具包括了自适应过程特征,能把行为建模技术的优点广泛应用到开发活动中。这些新特征类型能自动推动设计,它们不仅被用来定义设计的几何方面,而且还被用来定义象反射率、重量、金属材料或 NC 制造方法的行为和过程智能特征。通过把行

方和过程信息应用到它们的设计中去,工程师在开发过程的每一个阶段都能控制最终产品的质量、功能和成本效益。

2)使用的方便性:除了简洁的菜单描述和充分的两线帮助,其下拉式菜单使用户能以直观的方式进行各种操作和预先设定。Pro/Engineering 利用图形键将经常使用的命令和定制菜单联结起来,提高了使用速度。由于 Pro/Engineering 提供了在实体模型上进行草图设计的功能,零件的布置将变得容易和精确。

3)相关性:Pro/Engineering 是建立在一个统一的能在系统内部引起变化的数据结构的基础上,因此开发过程中某一处所发生的变化能够很快传递整个设计制造过程,以确保所有的零件和各个环节保持一致性和协调性。

4)参数化,面向零件的实体模型设计制作,Pro/Engineering 的零件设计功能是一些和系统内部蕴藏的知识,智能相联结的过程,可以画出非常复杂的几何外形。Pro/Engineering 设计的零件不仅包含制造工艺和成本等一些非几何的信息,而且还包括零件的位置信息以及他们之间的相互联系。这意味着在对零件进行布置时并不需要一个坐标系,零件自身知道他们是

如何和模型的其他部分相联系的。这就使得对模型的改动非常迅速,并且始终和最初的设计意图相一致。图附.1-2 为产品装配。

5)超大型部件:Pro/ENGINEER 的核心技术是以部件为中心,所以它能使工程师高效率地设计、归档和管理任意大小的产品部件。Pro/ENGINEER 利用新的专利技术包括简体技术,在精简图中准确表示曲面和物体属性数据。它可用来代替超大型部件中的复杂子部件,或用来获取反映移动部件行为的完整运动包络。由于精简包络模型并不包括内部的详细资料,所以,可以与供应链上的其他公司共享,且不泄露专有设计数据。

6)以 Internet 为中心:为了促进企业范围内的信息交流,提高桌面系统的访问能力,Pro/ENGINEER 提供了全面的以 Internet 为中心的工具。工程师和非技术用户可以使用许多易用、支持 VRML、HTML、URL 的工具,进行在线浏览、交互访问和共享 Pro/ENGINEER 的设计。而且,从 Web 网上还可以得到 PTC 公司新的 InPart,这是业界最大的在线零件目录,其中有来自领先部件供应商的成千上万个以前创立的 Pro/ENGINEER 标准设计。



图附.1-2 Pro/Engineering 的产品装配。

## 2.2 Pro/Engineering 功能

### (1) 零件设计

1)生成草图特征。包括凸台和凹槽,以及冲压的,旋转的,沿二维草图扫过的轨迹槽沟,或两个平行截面间拼合的槽沟。

2)生成“Pick and place”特征,如孔,通孔,倒角,圆角,花,规则图,法兰盘,棱等。

3)草图美化特征。

4)参考基准面,轴,点,曲线,坐标系,以及为生成非实体参考基准的图。

5)修改,删除,压缩,重定义和重排特征,以及只读



特征。

6)通过向系列表中增加尺寸生成表驱动零件。

7)通过生成零件尺寸和参数的关系获得设计意图。

8)产生工程信息。包括零件的质量特性、相交截面模型、参考尺寸。

9)在模型上生成几何拓扑关系的曲面的粗糙度。

10)在模型上给定密度、单位、材料特性或用户专用的质量特性。

11)可通过 Pro/FEATUER 增加功能。

## (2)装配设计

1)使用命令如 mate、align、insert 等安放组件和子装配功能生成整个产品的装配。

2)从一个装配中拆开装配的组件。

3)修改装配时安排的偏移。

4)生成和修改装配的基准面、坐标系和交叉截面。

5)修改装配模型中的零件尺寸。

6)产生工程信息、材料清单、参考尺寸和装配质量特性。

7)可通过 Pro/ASSEMBLY 增加功能。

## (3)设计文档(绘图)

1)生成多种类型的绘图视图。包括总图、投影图、附属图、详细图、分解图、局部图交叉截面域和透视图。

2)完成扩大视图的修改。包括视图比例和局部边界或详细视图的修改,增加投影图、交叉截面视图的箭头和生成快照视图。

3)用多模型生成绘图。从绘图中删除一个模型、对当前绘图模型设置和加强亮度。

4)用草图作参数绘图格式。

5)操作方式包括显示、擦除、开关视图、触发箭头、移动尺寸、文本或附加点。

6)修改尺寸值和数字数据。

7)生成显示、移动、擦除和用于标准注释的开关视图。

8)包括在绘图注释中已有的几何拓扑关系。

9)更新几何模型的组成设计的改变。

10)专门绘图的 IGES 文件。

11)标志绘图指示作更改。

12)通过 Pro/DETAIL 增加功能。

## (4)通用功能

1)数据库管理命令。

2)在层和显示层上放置零件的层控制。

3)用于距离的测量、几何信息角度、间隙和在零件间及装配的总干涉命令。

4)对于扫视、变焦距、旋转、阴影、重新定位模型和绘图的观察能力。

## 2.3 其他产品模块

(1)Pro/ASSEMBLY 是一个装配管理系统,它能提供用户用表格的手段生成一些装配的系列并可自动地在装配中更换零件。Pro/ASSEMBLY 中也包括 Pro/PROGRAM 模块,它提供了开发环境,运行参数化零件和装配设计的自动化程序。事实上,这些程序可使非技术性的用户产生自我定制的设计,而只需输入一些简单的参数。

(2)Pro/CABLING 提供一个全面的电缆布线功能,用户可以在 Pro/ENGINEER 的装配中布置和设计三维电缆和电线束。这些三维线束可在电器及机械部件设计和装配中并行进行。使用户可尽量优化他的设计以达到机械及电缆空间上的要求。

(3)Pro/CASTING。利用 Pro/CASTING,铸造和模具工程师可以很容易地设计铸模的几何形状、模具的容腔和型芯以及最终制造铸模,评估产品设计对制造方便性的影响,这可以加快模具和产品的生产设计过程,由于 Pro/CASTING 所具有的参数性,在产品的设计阶段铸模的设计就可以具有多次反馈循环,这能极大地优化制造工艺并缩短产品进入市场的时间。

(4)Pro/CAT 允许用户在 Pro/ENGINEER 和 CATIA 之间进行设计数据的交换,CATIA 模型可以直接输入到 Pro/ENGINEER,并可加上 Pro/ENGINEER 的特征和实现参数操作,而 Pro/ENGINEER 模型也可以输出到 CATIA 系统中。

(5)Pro/CDT 允许数据交换转换为 CAAM 绘图,从 Professional CADAM 到 Pro/ENGINEER 中。

(6)Pro/COMPOSITE。由对复合设计要求所提供的基于特征的工具自动进行设计、制造和复合材料文档生成。

(7)Pro/DESIGN 用于加速大型设计和复杂装配的自顶向下的设计过程,此模块工具包括概念设计的二维非参数化装配布置,用于概念分析的二维参数模型的布置和用于组件的三维布置模型。

(8)Pro/DETAIL 直接从 Pro/ENGINEER 实体模型生成全关联的 ANSI、ISO、DIN 或 JIS 标准绘图。Pro/DETAIL 还提供有效的尺寸标注公差和 Pro/ENGINEER 所固有的视图生成能力。

(9)Pro/DEVELOP 是一套程序开发工具,它把用户用 Pro/ENGINEER 开发的软件和用 Pro/ENGINEER 开发的第三家的应用软件以及定做的 Pro/ENGINEER 功能集成在一起。

(10)Pro/DIAGRAM 提供了广泛应用领域的原理图中的简图信息的生成工具,如电子线路、管道布置、HVAC 流程和计划管理等。

(11) Pro/DIEFACE 允许冲压模具的设计者定义用于钣金加工和其他深拉(deep drawn)加工中的复杂的部件。给用户提供了参数化模具曲面生成的可能,很容易改变和方便的优化钣金的可模锻性。

(12) Pro/ECAD 提供了 Pro/ENGINEER 和电器 CAD 系统之间的设计数据的交换,所交换的信息包括电路板的轮廓线、保持域和高度的限制范围,以及组件的形状和定位。Pro/CAD 将机械与电子工业设计链接起来,并允许并行地开发产品。

(13) Pro/FEATURE 允许用户生成一些自定义特征和复杂设计特征,如各种截面造型、壳体造型、三维扫描特征、各种由非平行交叉截面造型特征、薄墙设计、特征组以及其他很多特征。Pro/FEATURE 也可以拷贝和镜向特征。

(14) Pro/FEM-POST 提供了在集成环境中各种有名的有限元分析软件的全套的有限元分析后置处理程序的能力。这个紧密集成的环境允许在零件几何文本中获得图形和数字,并且提供用于自动驱动模型几何的能力。

(15) Pro/HARNES-MFG 允许用户生成加工数据,如钉固板和用于电线和电缆生产的附图布置,Pro/INTERFACE 提供了 Pro/ENGINEER 与其他设计自动化系统之间的各种标准数据交换格式。转换器支持由其他系统输入 IGES、SET、VDA、CGM、和 DXF,Pro/INTERFACE 也输出信息到这些相同的标准中,同时也提供 SLA 文件和 Neutral、Patran、Supertab 文件格式。

(16) Pro/LEGACY 可以使工程师在 Pro/ENGINEER 环境中对 2D 和 3D 数据文件进行维护,它的流畅的用户界面经过了优化处理,因此工程师们可以在 Pro/ENGINEER 环境中对模型的表面形状、线框等进行修改综合而不必重新生成。

(17) Pro/LIBRARYACCESS 提供常用的标准数据库的存取,包括标准零件和特征、加工刀具和夹具电缆设计的连接器、组件模块和电器、管道和加热符号等,用户可即时使用这些标准库而无需重新生成,这样可提高生产效率。

(18) Pro/MANUFACTURING 从加工模块中生产加工工艺计划、刀具编程和时间价格比的估算。Pro/MANUFACTURING 将工艺步骤与设计模块链接起来,这样,当设计改变时,加工信息可自动生成。此模块支持的加工过程包括钻孔加工车削加工和直到五轴的铣削加工。

(19) Pro/MESH 为有限元分析自动地准备实体模型,参数化和边界条件可以直接应用于实体模型,Pro/MESH 支持局部网格的精确表示和与标准分析

软件相兼容的输出格式。

(20) Pro/MOLDESIGN 用于模具部件和模具装配的设计,Pro/MOLDESIGN 包括浇注系统、浇道、浇槽等特征。并且允许用户做注模模拟。Pro/MOLDESIGN 为用户节省大量的时间,并使用户的零件设计和模具设计并行进行。

(21) Pro/NC-CHECK 提供一个图形工具,用于对铣削、车削、钻镗、仿形、冲压、冲裁加工所产生物料的模拟清除。

(22) Pro/PDM 为管理开发和支持大规格的参数化设计提供了一套综合工具。这套工具包括:设计小组并行工作的项目计划调度;中央文件和中间数据文件存储;存取控制;释放程序;ECR 跟踪;版本历史;产品结构管理;电子笼形线圈和广泛的查寻、浏览、报告工具。

(23) Pro/PIPING 是一个综合的管路设计软件包,在 Pro/ENGINEER 装配中具有三维管路的设计功能。三维管路设计与所有产品设计并行完成,允许机械工程师和液压工程师在没有物理模型时完全优化他们的设计。

(24) Pro/PLOT 提供支持工业标准的绘图仪的输出。包括 HPGL2、Versatec、Calcomp Gerber Photo Plotters。

(25) Pro/REPORT 为用户报告提供了功能强大的格式环境,如:材料表、工程改变顺序、属类表、和制造加工路线表等。

(26) Pro/REVIEW 允许用户察看所有 Pro/ENGINEER 数据,包括绘图、零件、装配、和制造信息。用户可以观看、评价和标记对象,但不能改变它们。Pro/REVIEW 对需要存取 Pro/ENGINEER 模型的所有成员去进行价格评估、察看设计和市场情况都有益处。

(27) Pro/SCAN-TOOLS 能使工程师光顾产品,能从干涉坐标扫描数据文件中产生曲面。Pro/SCAN-TOOLS 曲线和曲面形成工具允许工程师通过操作控制多边形去优化几何形状并实时观察改变的结果。

(28) Pro/SHEETMETAL 可进行钣金的设计和制造。用户可开发参数化钣金造型和装配。并可组合排样多种平面图案和指定加工程序。例如冲压、冲裁开槽等。

(29) Pro/STEP-INTERFACE 可以让用户按 ISO10303 标准 CSTEP 输入或输出数据。

(30) Pro/SURFACE 扩大 Pro/ENGINEER 的能力,以便解决复杂曲面的造型问题。它提供了很容易生成流线型表面的能力,例如样条驱动(Spline-driven)、多截面(Multi-section)和可变扫描(Variable sweeps),

以适应整个工业领域的要求。

## 2.4 Pro/E 2000i<sup>2</sup> 新功能

美国 PTC 公司在 2000 年 3 月份向广大用户发布了 PRO/ENGINEER 2000i<sup>2</sup>。在很多用户还没有把 PRO/ENGINEER 2000i 的行为建模等新技术深入了解的时候, PRO/ENGINEER 2000i<sup>2</sup> 又带着 440 多项改进和广大 PRO/E 用户见面了, 到底性能有哪些改进呢? 就让笔者把 PRO/ENGINEER 2000i<sup>2</sup> 的新功能简单介绍给大家, 不当之处, 还望批评指正。笔者认为 2000i<sup>2</sup> 和 2000i 相比, 功能要强大的多。因为 2000i<sup>2</sup> 和 2000i 相比, 有了更多人性化的设计, 其中包括可视化检索 (Visual Search)、形状索引 (Shape Indexing)、特征灵活性 (Feature Agility)、行为建模技术 (Behavior Modeling)、疲劳预测 (Fatigue Prediction)、CDRS 渲染 (CDRS Rendering) 等等。对于工程师来说, 这些人性化的设计可以让工程师把时间更多的用在优化设计和产品的创新上, 从而提高设计效率。

在这众多的新功能中, 最令人欣喜的就是可视化检索 (Visual Search) 了。如果您一直从事产品设计的话, 您一定有过这样的经历: 为了寻找一个其他工程师做好的零件或装配, 需要从几十甚至上百个零件、装配中去不停的打开或关闭一个又一个窗口。这样的工作让人厌烦, 并且效率极低。现在有了 PRO/ENGINEER 2000i<sup>2</sup> 的可视化检索 (Visual Search), 一切都简单多了。可视化检索功能就是让工程师在调用零件、装配图样时, 可以预览即将调用的零件。更让人惊讶的是, 您还可以和在 PRO/E 操作界面中一样用鼠标加 Ctrl 键去旋转放大或移动它去看更详细的结构。而完成一个超大型部件装配的预览只需要几秒钟的时间, 让人不敢相信。据 PTC 称, 这得益于 PTC 新的远程数据调动技术和专用多线程图形技术。在您打开预览功能的同时, 零件或装配的基本形状立即出现在您的眼前, 它只是一个近似的表示, 几秒钟后, 零件的细节越来越清晰。可视化检索具有智能化, 并且支持多线程处理。当您缩放模型时, PRO/ENGINEER 只会载入浏览区域内的部位的数据而忽略其它部位数据。您还可以在装配浏览窗口中直接选择子部件和零件, 来查看它们的结构。最后还有一点需要告诉大家, 那就是可视化检索支持 INTERNET, 您只要输入文件在 WEB 服务器上的位置, 就可以把有关该模型的图形化数据, 输送到预览窗口中。有了这项功能, 您就可以通过 INTERNET 实现协同工作了。既然支持 INTERNET, 局域网就更不在话下了, 如果您是项目负责人, 您就可以通过在打开窗口里切换到网上邻居的方式, 看看大家是不是都在按你的计划工作。可视化检索不仅给工程师带来了

极大的便利, 也为浏览庞大的产品数据库带来了方便。

形状索引 (Shape Indexing) 也是 PTC 正在申请专利的先进技术之一。它利用智能过滤技术, 把当前的模型与数据库中的已有模型进行比较, 找出那些具有相似几何体的模型, 以提高设计的重用性。充分利用这项功能, 可以使企业减少许多重复设计的工作, 也为数据的管理带来方便。对于交通产品, 机械制造业等等来说意义重大, 因为他们制造的许多产品都使用相似或相同的部件。有了这项功能, 也许在新产品的开发中, 您只要投入一半的工程技术人员就可以, 另外一些部件只是需要调出来略微修改就可以用于新产品了。PTC 称, 由于利用了自动过滤和快速过滤技术, 如果对一个包含 300000 个模型的企业数据库进行检索, 只需要 2 秒就可以得到结果。这样的速度恐怕连挑剔的用户也会无话可说。

特征灵活 (Feature Agility) 也是这次 PTC 大力宣传的技术热点。因为在当今的产品设计中, 为了满足客户的要求, 适应市场的变化, 更改设计是司空见惯的事。而如何缩短更改周期就成了需要研究的问题。在 PRO/ENGINEER 2000i<sup>2</sup> 中, 增加了一种称为意向参考的新功能。使用这种功能, 设计者可以定义特征建立的方式, 利用这种定义方式, 你可以灵活的构建特征, 并且在特征发生更改时仍然保持你的设计意图, 从而把设计人员从烦琐的设计更改中解放出来。在 PRO/ENGINEER 的用户中, 经常会遇到这样的问题: 当你试图修改一个特征时 (特别是这个特征是较早构建的), 经常会跳出参考丢失, 后续特征不能构建的对话框。你需要手工重新定义每一个失败的特征直到零件完全再生成功。这种重复性的工作使设计人员不能全力投入设计, 而必须不断考虑这类特征的不稳定性。利用意向参考, 工程师可以更方便的指定特征建立的方式。比如: 当工程师需要在一个突起的柱体和其它特征的相交线上创建圆角时, 可以不参考特定的边界来建立圆角, 而参考需要导圆角的突起柱体与其它特征的相交边界来建立圆角, 这样特征就不会依赖于特定的边界而产生不稳定性。当工程师把柱体的形状和位置移动时, 圆角特征会自动适应这一变化, 这样就会提高设计的灵活性, 工程师也更轻而易举的完成设计的更改。这是 PTC 的柔性工程所带来的变化, 这种贴近用户的全面而细致的考虑是三维工程软件的发展趋势。

行为建模技术是从 PRO/ENGINEER 2000i 开始推广的新技术。这种行为建模技术被业界作为第五代 CAD 技术。它通过把导出值 (比如容积) 包含到参数特征中, 再反过来使用它们生成和控制其它模型的几何图形。使用行为建模技术, 用户首先要定义一个工程分析模型, 其中包括名称、类型和定义。接着, 他们要建立

“操作”定义和引用一类用在分析模型中的新特征,比如体积,然后为分析模型设置约束条件-包括目标值、一个参数的最小和最大值。系统会出现解决方案的图形,协助用户为设计选择最优方案,利用行为建模技术的自动求解能在最短的时间内,找到能满足工程标准的最佳设计。相对于传统的手工反复操作的方法,行为建模的全自动处理有很多优点。对于简单的问题可以方便的求解,复杂的问题则可以在很短的时间内完成寻找解决方案的任务。有了这个强大的功能,产品的设计创新才会更快更好。

疲劳预测(Fatigue Prediction)是 PRO/MECHANICA 的最新功能,它可以提供工程师预测和改善设计的疲劳性能,可以帮助设计人员在产品还没有生产之前就识别会缩短产品寿命的潜在故障。为企业在设计阶段的产品质量提高带来了极大的方便。CDRS 渲染(CDRS Rendering)增加了一些新功能如允许纹理处理、改进的照明、雾化、镜头闪光、景深等效果。其它方面如模具设计、数据交换、曲面处理、高速加工等等也进行了很多改进。

总之,PRO/ENGINEER 2000i<sup>2</sup> 的多项改进使 PTC 在大型三维工程软件业界进一步奠定了龙头的地位。在很多软件争相向参数化靠拢的时候,PTC 又推出了更新的概念如行为建模技术、柔性工程技术等等,这些贴近用户的设计必将带来更大的市场,势必会推动 PTC 等公司研发更新的技术,就让我们拭目以待吧。

## 2.5 Pro/ENGINEER 野火版(Wildfire)

Wildfire 版可以让您通过互联网来连接 Windchill 应用,以实现真正的设计共享,即实时连接和访问互联网(或您的内部网)上的产品信息。

借助像菜单映射器(Menu Mapper)这样的特征功能,使用“从旧到新”命令并获得大量的帮助工具,对于老用户来说,移植将易如反掌。Pro/ENGINEER Wildfire 比以前更易用。

### (1) 产品功能

- 1) 利用直接建模交互技术来更快速地进行设计。
- 2) 使用 WARP 工具,对设计进行变形、扭曲、弯曲和拉伸操作。
- 3) 借助仪表板功能,来体验直观的工作流。
  - 1) 借助人-与人的连通性进行交互,来完成动态设计评审。
  - 2) 通过与 Windchill 的无缝集成,来访问产品和项目信息。

### (2) 客户得到的好处

- 1) 可以提高生产和过程效率。

- 2) 借助直接建模环境,创建一流设计。

- 3) 拥有与 Pro/ENGINEER 以前版本进行互操作的功能。

- 4) 可以简化许可管理和安装。

### (3) Pro/ENGINEER 设计

技术创新和效率优势,使 Pro/ENGINEER 成为产品开发最健壮的万用参数化实体建模解决方案。

### (4) Pro/ENGINEER 仿真

我们的 Pro/ENGINEER 仿真软件,可以让工程师对设计进行结构、动力学、热传导和耐用性等性能测试,从而进行优化。它需要最少的物理原型,可以提高创造力,并有助于在更短的时间内交付更优秀的产品。

### (5) Pro/ENGINEER 生产

通过满足型腔模和成型模行业的特种自动应用,以及适用于 NC 编程的强大而高效的 CAM 解决方案,可以提高生产率。

### (6) Pro/ENGINEER 布线系统

可以帮助设计师和工程师更快、更简单地建立复杂的互连布线系统图,它是整个产品开发过程必不可少的一部分。它能让用户建立配线图、P&ID 图、方块图和原理图。

### (7) Pro/ENGINEER 工作组数据管理

Pro/ENGINEER 工作组数据管理可以让您的小组建立一个安全、可伸缩的环境(并在其中工作),它能提高设计人员的效率,减少设计人员之间的争论。为了满足我们独特客户的这些特定需求,您可以从这三种部署中进行选择。

## 3 I-DEAS 软件

### 3.1 I-DEAS 软件概述

SDRC 公司的 I-DEAS Master Series 在一个电子环境中集成了建模、仿真、优化、加工和测试的所有产品开发所需的功能。I-DEAS 提供了从设计的最先进技术 VGX,到全面仿真和分析能力,直至复杂的 7 轴连动、5 轴加工能力,使 I-DEAS 成为所有机械设计自动化软件中功能最强的软件系统。

I-DEAS 为用户展现了一个动态的高度集成的开发环境,帮助产品开发队伍真正以团队方式来协同工作,使团队成员之间更方便地交换信息,在协同环境中快速共享在整个产品研发过程中所做的工作。

I-DEAS 与 SDRC 自身的 Metaphase 这一当今最先进的产品数据管理(PDM)软件的无缝集成为企业提供了掌控产品开发全过程的保证。I-DEAS 还允许设计团队基于公共主模型同时展开工作,由此生成的

数字样机可提供以前只能靠物理样件实验才能得到的答案。

该软件的特点有:

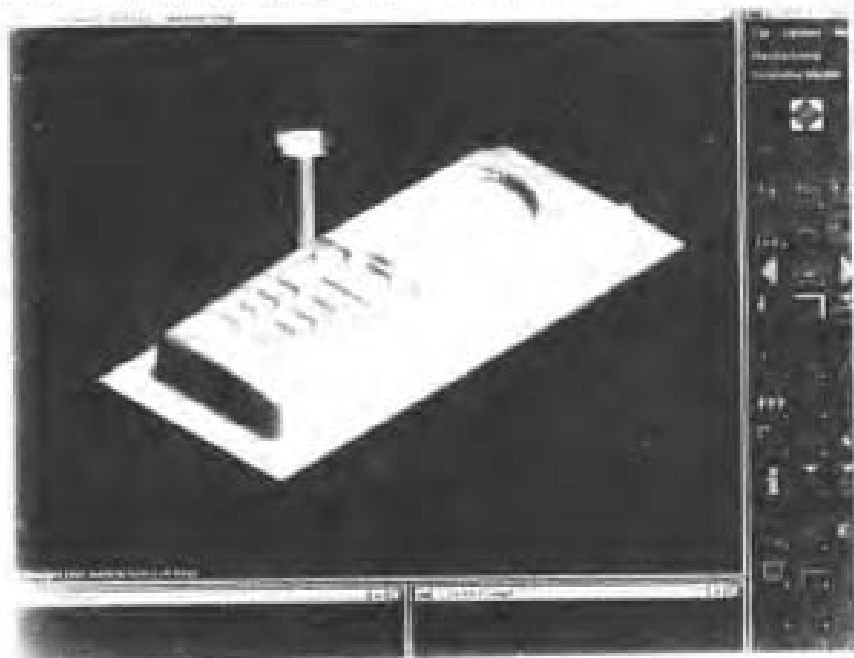
#### (1) 先进的 VGX 变量化技术

VGX 将直接几何描述和历史树描述这两种目前最好的造型技术创造性地结合起来,它解放您的思想,开拓您的思路,使您集中精力于创造性概念设计上,而不是在繁琐的软件使用机制上。

许多 CAD/CAM/CAE 系统的多层次菜单功能导致了复杂的命令结构,而 I-DEAS 特别易于使用,它既

功能强大,又操作简单。

作为 I-DEAS 核心的 VGX 技术提供了动态引导器(Dynamic Navigator)这样独特的关键技术。作为一个十分好用的交互工具,动态引导器可自动识别并预增强装配体、零件、边、(面)面、线框、草图、单个几何实体和所有约束,参与用户的下一步操作,它具有直接在实体零件上任意位置勾画草图的能力。有了 VGX,您可以直接在三维数字模型上进行增、删、改任一或一组特征的操作,既直观且随意,这时的产品设计和修改就变得如同手捏面团一般(见图附-1-3)。



图附-1-1 I-DEAS 的模具加工

#### (2) 实时协同工程

真正的团队协同环境极大地改进了设计过程和最终产品的质量。I-DEAS 为产品开发队伍的所有成员(尽管技术水平不同)提供了“并行相关性”(Concurrent Associativity),使他们在开发全过程中均能保持联系,以确保永远使用正确版本的数据。I-DEAS 提供了一个可存入和调出零件、装配件、工程图和特征的数据库,并且能自动跟踪版本的更新。利用 I-DEAS 提供的 WEB 工具,团队成员可快速且简单地生成、修改、管理和共享数据,而不必去关心它们的物理位置。

I-DEAS 无缝地集成了多个应用领域,允许在一个单一的电子环境中生成、仿真、优化、建档、制造和测试您的产品设计概念。另外,由于 I-DEAS 允许您同时使用其他商业应用软件,或从其他软件迁移过来,这样,您以前的投资得到了有效保护。

I-DEAS Team Data Manager 是与 I-DEAS 紧密集成的团队级 PDM 软件。它为有权限的团队人员提

供了对主模型数据的访问,这样不仅使数据访问和共享更方便,也能防止零件被不同的用户进行不可控制的修改。通过 I-DEAS/Metaphase 接口,数据能无缝升级至企业级 PDM 系统。通过这个接口,I-DEAS 的产品数据,用户专有程序的数据以及其他商用系统的数据均可以由 SDRC 公司的 Metaphase Enterprise 软件统一管理。

#### (3) 功能高度集成

I-DEAS Master Series 集工程领域所需功能之大成,为您提供从产品造型、产品仿真、产品测试直至产品加工的全部功能。诸如结构分析、热力分析、优化设计、耐久性、测试数据的采集处理及声音品质分析等真正提高产品性能的必备功能也一并包含在内。

#### (4) 良好的开放性

I-DEAS 可兼收并蓄其他商业或自用软件,实现与电子、机械设计、分析、测试、加工、快速成型以及其他具有并行工程功能的应用软件的数据共享,保护您以前的投资。I-DEAS 基于 CORBA 开放标准, Open I-

DEAS 提供一致的、面向对象的 Client/Server 界面。此外,任何标准的 WEB 浏览器都可以查看 I-DEAS 以 VRML 超文本格式生成的模型文件。

### 3.2 I-DEAS 软件功能

#### 3.2.1 产品设计(Core Master Modeler)

##### (1) 建模(Master Modeler)

I-DEAS 软件的产品建模是基于特征的变量化实体造型系统。在 I-DEAS Master Series 中,无论是单个模块还是成组应用都是以 Core Master Modeler 为核心来运行。这种先进的核心实体系统结构决定了它必然是以主模型(Master Modeler)为单一数据库,并且所有的模块之间的数据全部并行关联。

1) 完全一体化的变量化设计环境。支持统一的线框、裁剪曲面和实体造型。

2) 先进的超变量几何(VGX Variational Geometry Xtended)技术。既可进行全几何约束的参数化设计,又可进行任意几何与工程约束的自由创新设计。

a. VGX 技术贯穿于 2 维草图设计、3 维零件造型、直至装配体设计全过程。

b. 提供变量化草绘、建立变量方程、设计变量特征的能力。

c. 直接修改与基于设计历程修改相结合。可随时灵活地修改原约束中、建入新约束和删除旧约束,或进行与造型顺序无关的尺寸标注,而不必关心设计顺序;可在欠约束情况下进行创新设计。

d. 可在复杂的曲面边缘上自动生成凸缘。

3) 基于特征的设计。提供特征库(预定义特征、表驱动特征和用户自定义特征)、特征操作和特征管理(历史图形树浏览、特征重排序、特征删除、特征消隐等)能力。

4) 多种复杂形体设计与修改能力。

a. 复杂几何形状设计与修改。

b. 自动隐藏线消除。

c. 物性的直接计算。

5) 独创的集产品几何信息(变量特征、设计历程、工程约束方程)、工艺信息(尺寸、坐标系、公差配合、形位公差、材料及物理属性)和加工工艺信息等为一体的完整的信息主模型。

6) 一体化的项目组级产品数据管理环境。通过 IDM(项目数据管理),TDM(团队数据管理)和并行相关性(concurrent Associative)进行全过程统一数据管理,支持团队协作工程。

7) 基于独创的动态引导器(Dynamic Navigator)

和动态图表的最易学易用的界面,可随时捕捉设计者意图,用预增亮和符号提示各种尺寸和位置关系,操作简单直观,为设计者准备下一步骤所用工具;引导设计过程,使设计者集中精力于设计思路与创新。

(2) I-DEAS Web 数据访问(I-DEAS Web Access)——经由 Web 直接访问 I-DEAS 数据

1) 强化 TDM(团队数据管理)功能。无论您是否在 I-DEAS 环境下工作,无论您是否在工作现场,您均可通过 I-DEAS Web 数据访问功能进入 TDM 系统,为设计添加属性或文件。

2) 通过 Internet 直接浏览装配,获取最新的设计版本。

3) 通过标准格式交换数据,获取数据并进行数据管理。

4) 无须经由其他绘图管理系统,便可直接访问 I-DEAS TDM 查看或绘制图样。

5) 由于本模块基于 JAVA 技术编写,因此可直接用于 Internet 和 Intranet,易于对软件进行维护、管理及二次开发。

#### 3.2.2 产品装配(Assembly Set)

##### (1) 装配(Master Assembly)

I-DEAS 软件的装配设计可以配置、设计和管理大型机械装配体的多用户并行工作环境。

1) 支持大尺寸、无限数量零件和多层次超大装配结构。

2) 多种灵活的零部件定位方式和约束关系,确保一个零部件的修改会自动更新其他所有相关零部件。同时可动态、实时地显示装配的自由度状态。

3) 团队环境中的从自顶向下到自底向上的设计。总设计师确定装配结构及总体设计后,多个设计者同时进行零部件细化设计,可随时自动更新总体设计。最后总设计师进行装配件的干涉检查和运动仿真等。

4) 设计评估。含干涉检查、运动学和动力学仿真与分析、装配体误差分析及物性计算等。

##### (2) 机构设计(Mechanism Design)

I-DEAS 软件的机构设计可以模拟机构的复杂运动,并进行动力学分析

1) 利用图标显示的各零件间关系的约束链,建立或修改装配关系。

2) 局部和全程运动干涉检查。

3) 多铰链、多环机构设计。

4) 运动学分析——计算相对位置、速度和加速度。

5) 系统可进行“大”位移的动力学分析。如内置动力学解算功能;碰撞;在外力作用下的机构动力学响应

分析。

6)与 ADAMS 软件的专用接口,以定义更复杂机构及加载和解决更高级动力学仿真问题。

### (3)装配公差分析(Tolerance Analysis)

I-DEAS 软件的装配公差分析是装配体的三维统计公差分析软件。

1)建立三维产品的几何模型,公差装配过程和装配测量的虚拟样件以模拟装配加工过程并预测机械系统装配公差所产生的影响。

2)提供数千种装配测量方式,模拟装配体公差的分布及每个尺寸公差的影响。

3)用形象化图形界面显示各零件运动到通常位置,最大位置和最小位置状态并计算公差统计分布,经与设计规格比较分析误差原因,并迅速修改误差或装配过程。

### 3.2.3 曲面设计(Surfacing Set)

#### (1)曲面造型(Master Surfacing)

I-DEAS 软件的曲面造型是基于 NURBS 的复杂雕塑曲面设计系统。

1)双精度 NURBS 的多种精确曲面造型方法,并与线框、实体完全集成。

2)复杂曲面造型及全局或局部曲面控制能力。如通过修改尺寸或几何约束的变量化形状控制;变量方程控制的曲面修改;切矢控制的曲面修改。

3)变量扫描,支持多轨迹扫描、变半经过渡面及尺寸驱动的自由曲面。

4)基于最小能量原理的形状控制技术,可实现对曲面的高级操作,如推、扭、凹、抽取等。

5)曲面分析工具。分析曲面上任意处曲率、高斯等参线、平均曲率、主曲率、切矢、法矢,检查曲面拼接。

6)与其他软件的曲面数据转换。

#### (2)数据转换

1)3D IGES 数据转换器(3D IGES Data Translator)。

2)VDA—FS 数据转换器(VDA—FS Data Translator)。

3)SET数据转换器(SET Data Translator)。

4)快速成型机数据转换器(Rapid Prototyping Data Translator)。

### 3.2.4 产品绘图(Core Drafting)

(1)二维绘图(Drafting)是高效的产品绘图系统,支持汉字及图标

1)与 I-DEAS 的 Master Modeler、Master Surface 或 Master Assembly 等模块集成,将 3D 上模型一切信息

转为各种 2D 工程图,并可往 3D 模型上添加设计细节。

2)3D 与 2D 的双向信息关联。

3)独立且强大的 2D 绘图环境。支持变量化 2D 绘图和 2D 概念设计。

4)支持多种尺寸与公差标准及双向数据转换。

5)自动生成产品 BOM 表,并标注装配序号,同时还可生成工程图的管理。

(2)审图与标注(View & Markup)解决图纸审查及修改

1)提供审查和标记工程图能力,并能自动更新工程图。

2)标记方式有删除修改、移动、尺寸修改、公差检查、文字注解等。

3)支持 Drafting 图形、图片或 CGM 格式文件。

4)支持多人并行工作方式,并与原设计图双向沟通。

### 3.2.5 加工系列(Manufacturing Set)

#### (1)创成式加工(Generative Machining)

可以在一个统一的环境中完成加工工艺计划、工具定义和 NC 编程任务。

1)由于包括了完整的加工环境(零件、毛坯、机床定义和夹具),生成的刀轨能自动避让零件的所有要素以及夹具等要素,保证刀轨反映加工真实情况,提高编程可靠性。

2)由于将毛坯集合进加工环境中,可以清楚地反映每一步加工毛坯的剩余量,并能自动地优化刀轨,根据每一步毛坯的余量加工,节省了不必要的空刀,提高加工效率。

3)面向团队加工。基于 I-DEAS 提供的强大框架,可以实现设计和加工的同步工程。可以在设计进行到一定阶段后开始加工工艺编程,这样可以保证零件的工艺性,缩短产品开发周期,当设计修改后,能自动反映到加工中来并随之更新刀轨。

4)创成式加工。能够将成熟的加工工艺内存为加工规则和方法,并在同类加工中调用。这样可以实现标准化加工,提高效率。

5)基于实体、自由曲面和特征驱动加工。由于采用了强大的算法,可以加工实体和自由曲面。对于任何数据转换得到的模型不必缝合成实体,直接加工,大大扩展了应用范围。

6)完整的三轴联动,5 轴定位铣削加工。如提供从粗加工到精加工及清模加工、刻字加工等多种走刀方式;多种加工选项,包括合理的进刀、运刀方式、多种切削方式和完备的加工选项。

7)支持多角二轴车床加工。如完整的车削刀具定

义;提供粗加工—精加工多种走刀方式;提供螺纹、切槽、切断等完备的加工方式。

8)5轴钻孔加工。如完整的钻孔加工方式;3~5轴钻孔一次完成。

#### (2)标准后处理

为任意数控机床生成标准的客户化的后处理程序,支持2~6轴车削加工;2.5~5轴联动铣削加工;文本化后处理文件,可以方便地编辑,实现客户化。

### 3.2.6 高级加工系列(Advanced Manufacturing Set)

#### (1)CAMAND建模(CAMAND Modeler)

强大的曲面和线框造型能力,快速生成加工模型。进行NURBS曲面造型(曲面拼接、剪裁、扫描、直纹面)与编辑(旋转、过渡、放样、分割)

#### (2)CAMAND三轴加工(CAMANDM Machinist)

提供3轴联动,4~5轴定位的NC/CNC高级曲面加工方案。兼备线框和表面造型能力,可以直接进行高级的图示化刀具轨迹编程,数控编程和机床刀具轨迹仿真。

1)多种类型加工(粗加工、平行直线运动、扫描平面加工、投射加工、边界精加工)。

2)多个曲面加工——以上加工类型均可在任意多个裁剪面上产生刀具轨迹。

3)进行后处理,产生指令文件。

#### (3)CAMAND多轴加工(CAMANDM Multax)

提供5轴联动CNC刀具轨迹、曲面造型与图示化工具。可进行3轴多表面加工和多达11种方式进行5轴加工。

#### (4)CAMAND车削系列(CAMAND Lathe)

1)提供2~4轴车削加工,含车螺纹、刮面、开槽、钻孔等。

2)仿真——模拟车刀、转台和零件的运动。

3)高度的集成性。

#### (5)CAMAND电加工(CAMAND Wire EDM)

生成2~5轴线切割刀具轨迹,以适应多种不同的线切割机床和EDM几何形状。

### 3.2.7 产品仿真(Simulation Modeling Set)

#### (1)仿真建模(Simulation Modeling)

有限元建模(FEM)

1)直接从I-DEAS零件主模型、装配体主模型或从其他CAD系统模型自动生成有限元网格。

2)单元库提供约60种单元,多种有限元建模方式。

3)多种载荷与边界条件定义方式。

4)提供丰富的图形和数据工具用于检查有限元模型的正确性和网格优化,分析结果的观察和后处理。

5)与20多个专业有限元分析软件的集成。

高级FEM(Extended FEM)——用于非线性有限元分析,提供自由网格、映射网格、人工网格等多种网格划分功能

梁建模(Beam Modeling)——专门用于各种梁的有限元建模,任意梁横截面造型、分析及后处理。

#### (2)仿真解算(Simulation Solution Set)

仿真顾问(Simulation Advisor)指导用户完成有限元建模、求解和计算结果解释的内置专家系统。

线性求解器(Model Solution Linear)用于线性结构热场和流场有限元分析的模块,以多种固定形式给出求解参数和输出结果,使运行简化。

模型响应(Model Response)通过线性求解器的扩展,预测有限元模型的动态响应能力。

优化(Optimization)用分析结果改进设计,包括设计目标和几何形状的优化,各种物理属性和材料的优化。

变量化分析(Variational Analysis)一次迭代(一次网格划分和一次解算)计算,变量化分析即可提供针对相应设计参数的重复变化的分析结果。实现一次解算,便得到多种方案的最佳效果。

#### (3)非线性求解器

非线性求解器(Model Solution Non-Linear)用于载荷、材料特性、接触条件或结构刚度随位移变化的非线性条件下的有限元分析。非线性求解器支持几何非线性、材料非线性、面面接触及它们综合的分析;还支持加载自动步长控制和弯曲分析等复杂问题。

### 3.2.8 钣金设计(Sheet Metal Design)

钣金设计可以进行钣金零件的设计与排样。

1)自动对变量化实体模型施加用户定义的弯曲度、应力释放和收缩补偿。

2)提供标准钣金特征库和钣金零件库。

3)自动转换折弯与展开状态,并协调两种状态下的设计细节及与装配件的还原关系

4)自动生成工程图和展开图、以及加工刀位轨迹。

### 3.2.9 电缆布线设计(Harness Design)

电缆布线设计提供装配环境下电缆线索的布线设计和管理。

1)提供装配环境下电缆线束的布线设计与管理。

2)变量化形状设计,便于快速进行定位器之间的线束定位,自动计算线束直径和电缆长度。

3)将配电数据和相应的机械连接线动态地集成,



自动检查缆线的干涉,约束其长度、布置线束路径。

4)实时渲染显示和动态修正布线。

### 3.2.10 管道设计(Sumpipe)

用以进行管道系统设计与结构优化。

1)可在管道系统结构中定义管道约束,定义管道零件(法兰“T”/“+”交叉,阀门等)定向管道零件、连接于装配、修改标准件、移动、旋转、删除任意管道零部件。

2)装配中已使用过的零件都存放在现有零件库里,以节省硬盘空间。

3)如果用户需生成一个新的标准件,该系统可在现有零件库中自动检索,如果有此件可被直接使用,如果没有该系统可在通用零件库中生成一个零件,根据用户需要的尺寸定义。零件一旦生成便存入现有零件库,同时自动进入正在使用的装配中。

### 3.2.11 注塑模架库设计(Moldbase Design)

根据标准模具架库和部件自动生成全套模架总成。尺寸驱动的智能模架特征可加到模架装配中,使用独特的三维动态预览功能调整特征的尺寸、位置。

1)标准部件的图视化功能可使确定模架尺寸的工作变得更加容易。

2)自动生成模架的分解图。

3)从十几种零件类型中选取合适的零件添加到模架中。每种类型的零件可用预览图形的方法进行定位。

4)定位销与模架相互关联,并随之进行变化。

5)根据需要随时改变模架几何尺寸,相应部件结构也会重新生成。

### 3.2.12 复合铺层设计(Laminate Composites)

用来进行复合层压结构设计、评估和显示。

1)提供层片特性(可由外部输入或内部计算而得)与复合层压结构的特性定义。

2)提供各种载荷分析。

3)复合层压结构附加于有限元模型的各种位置方式。

4)与 NASTRAN、ANSYS、ABAQUS 的接口。

### 3.2.13 电子系统冷却仿真(Electronic System Cooling)

电子系统冷却仿真用于进行电子系统中三维流场和热特性模拟。

1)与 I-DEAS 完全集成,直接通过接口使同一几何模型进行系统和单个元件的模拟。

2)可进行单个元器件多芯片组、散热片、线路板或完整电子系统的造型。

3)可优化元件的位置;预测风扇运行;探讨自然对流和强制对流的制冷方案;设置风扇和通风口的位置与尺寸;优化散热片的形状和尺寸以加大对流;通过热路和热传播进行热传导。

4)为与机械设计的集成提供了设计引导,并可建立复杂几何体的分析模型(含自动生成自由网格、控制网格密度、局部细化网格等);设计变更后网格的及时更新。

5)将强大的计算流体力学(CFD)技术与热力学技术结合为液体流动建模,模拟曲面上复杂的对流、传导、幅射及热特性。

6)模拟结果可用 3D 矢量、粒子路径、轮廓、准则等图形、表格或报告给出。它反映流体速度,风扇及通风口,流体及实体的流动温度、热度,流动压力、传导系数或湍流数据等。

### 3.2.14 复杂热交换仿真(TMG)

1)使用先进的有限差分控制体技术对非线性和瞬态热传导过程(含传导幅射、自由或强制对流,流体流动和相位变化等)进行快速有效的造型和分析。

2)与 I-DEAS 的完全集成性。可直接使用有限元模型并将热分析结果返回载荷结构模型。

3)有限差分热模型支持实体、壳、梁及其组合,任意自由网格,任意元素形状,及随温度变化的材料。

4)先进的仿真技术可模拟漫反射、镜面反射和透射,既可直接在零件表面定义幅射模型,也可连接装配体中各零件之间不相接的网格进行热幅射分析。

5)求解器运用多种控制方法和运行选项进行稳态或瞬态热分析。

### 3.2.15 注塑冷却仿真(Moldflow)

注塑冷却仿真是世界上最普遍使用的对注塑、冷却、模具和环境的交互仿真系统。

1)模拟浇注模具过程中塑料液体的流动。使用户可优化零件与模具的设计、材料和过程条件,以达到质量、成本和时间的最优平衡。

2)快速注塑填充分析,以确定模具填充中聚合物的特性,保证同样压力下液体同时到达模具的各个极限位置。

3)注塑冷却分析,用以优化模具和冷却回路的设计。

4)3D 热传播分析,用边界元素法确定热传播速率和温度分布。并可根据给定条件,分别采用自动分析和手工分析法。

5)提供 4000 种材料数据库和各种模具特征(插入、分模面、浇口等。)

### 3.2.16 产品寿命预测 (Durability and Dynamics)

1) 提供一组预测产品响应、安全系数和寿命的分析工具。

2) 当不存在低频动态模式或零件固有频率大大高于所有激振时,采用静态分析去评估响应和有限元模型的寿命。典型实例为同时承受周期载荷和热载荷的汽车发动机的寿命评估。

3) 当存在低频模式或结构的弹性振动能影响产品运行时,采用动态寿命分析,此时使用模式加速度方法生成瞬时动态响应。典型实例为承受随机路况的汽车车身的寿命分析。

4) 此模块计算强度安全系数和疲劳安全系数。根据力、转矩、加速度和位移评估动态响应;根据各种载荷评价寿命或损坏情况以进行疲劳评估。

### 3.2.17 推动声场仿真 (Vibro-Acoustics & Rayon Solvers)

是与结构分析和测试集成的声仿真软件。

1) 提供预测声音如何影响机械元件的结构性能及振动结构如何产生噪声的能力。

2) 与 I-DEAS 完全集成的推动声场仿真,执行推动声音建模,声场和材料特性及边界条件的定义,推动声场数据准备,结构和声场分析及动态测试。

3) 先进的求解器支持边界元, Rayon 体有限元求解器,分别解决内部腔体和外部声介质的声场问题。

4) 振动声曲面和声场结果包括结构加速度,结构均方速度,声压,声加速度和声强,及辐射能,吸收能,动能,形变能等多种能量与流体,并以集成图和基于几何的图形界面显示出来。

### 3.2.18 材料特性系统 (Material Data System)

1) 提供丰富的材料库来确定、修改和显示机械产品设计/仿真/加工用材料的机械/热/流动特性以及相关信息。

2) 用户可中 Form 组交互对材料数据进行查找、修改、列表、比较和管理。

3) 材料数据可为常数,多维表或方程(多达 5 个独立变量的方程)

4) 广泛用于结构分析(弹性、塑性、蠕变、粘塑性等),非结构分析(如作为温度与压力函数的空气密度),以及输出给 NASTRAN, ANSYS, ABAQUS 等系统使用。

## 4 CATIA 软件

### 4.1 CATIA 软件概述

CATIA 是由法国著名飞机制造公司 Dassault 开发并由 IBM 公司负责销售的 CAD/CAM/CAE/PDM 应用系统, CATIA 起源于航空工业,其最大的标志客户即美国波音公司,波音公司通过 CATIA 建立起了一整套无纸飞机生产系统,取得了重大的成功。

围绕数字化产品和电子商务集成概念进行系统结构设计的 CATIA V5 版本,可为数字化企业建立一个针对产品整个开发过程的工作环境,在这个环境中,可以对产品开发过程的各个方面进行仿真,并能够实现工程人员和非工程人员之间的电子通信。产品整个开发过程包括概念设计,详细设计,工程分析,成品定义和制造乃至成品在整个生命周期中的使用和维护(图附-1-4)。



图附-1-4 CATIA 的产品设计

作为世界领先的 CAD/CAM 软件, CATIA 可以帮助用户完成大到飞机小到螺丝刀的设计及制造,它提供了完备的设计能力,从 2D 到 3D 到技术指标建模,同时,作为一个完全集成化的软件系统, CATIA 将机械设计、工程分析及仿真和加工等功能有机地结合,为用户提供严密的无纸工作环境从而达到缩短设计生产时间、提高加工质量及降低费用的效果。

### 4.2 CATIA 软件功能

#### 4.2.1 机械设计

(1) 绘图-空间 (2D/3D) 集成 (Draw—Space (2D/3D) Integration)

CATIA 绘图-空间 (2D/3D) 集成产品将 CATIA 绘图和三维造型工作环境完全集成在一起。设计师和绘图员可以通过 3D 几何的投影和平面剖切直接生成 2D 工程图样。系统可从 3D 数据自动生成绘图视图和剖视图,并且用户可控制模型之间 2D 和 3D 的关联

性。绘图员可以结合使用 CATIA 的用于二维工程绘图的工程绘图(Drafting)产品以及 2D 线架和标注(2D Wireframe & Annotation)等其他产品,更可以充分体现该产品的优势。

#### (2) 工程绘图(Drafting)

CATIA 工程绘图(Drafting)产品是 2D 线架和标注(2D Wireframe & Annotation)产品的扩展。工程绘图(Drafting)产品可以允许绘图员更方便地生成工程图样。该产品提供了全套的绘图工具,可用于文本注释、尺寸标注、客户化绘图标准、2D 参数化设计以及 2.5D 视图定义。

#### (3) 2D 线架和标注(2D Wireframe and Annotation)

CATIA 2D 线架和标注产品为快速方便地建立、分析和修改 2D 几何及对 2D 和 3D 的 CATIA 模型进行显示提供了工具。作为降低费用的一种手段,该产品允许建立基本的几何图形和标注。若需要全尺寸标注功能,CATIA 绘图产品可为制图提供全面的功能。

#### (4) CATIA—CADAM 工程绘图(CATIA—CADAM Drafting)

CATIA—CADAM 工程绘图是一个高效、易用的用于设计和工程绘图应用的产品。在其基本产品中包括有:2D 和 2.5D 设计和工程绘图,包括有自动或人工指定的参数、几何和约束的变量设计功能;Sketch it 一个全新的革命性功能,可在用户进行徒手草绘图形的同时,动态地建立起全约束的参数化几何。

#### (5) 动态草图器(Dynamic Sketcher)

CATIA 动态草图器可作为一个工具在用户快速草绘 2D 轮廓图并将其作为 3D 设计的一部分时使用。用户可以直接操作共面实体几何并且编辑形状、改变自由度或修改相关约束。动态草图器在 2D 窗口中不断地分析几何以检查约束冲突。通过对轮廓几何约束的自动记忆,该产品可以捕捉到设计意图。用户也可配置约束参数,并且对约束进行增加、显示和编辑。

#### (6) 3D 线框(3D Wireframe)

CATIA 3D 线框是 CATIA 系统几何造型的关键产品,据此用户可以精确地建立、修改并分析 3D 几何。3D 线框产品对构造在许多设计系统中并不常见的 3D 几何提供了广泛的方法。对需要一个基于具有逻辑可扩展的 CATIA 的入门 CAD 系统的用户来说,该产品可作为一个独立的解决方案。3D 线框对支持象设计、分析或制造这样的配置提供了通用工具。

#### (7) 精确实体造型(Exact Solid3)

CATIA 精确实体造型可用于完成数控加工和有限元分析等后续应用所需的实体模型定义。丰富的分析工具和结构实体几何(CSG)技术可大大方便实体的

定义,并可优化设计。

#### (8) 3D 参数化变量造型器(3D Parametric Variational Modeler)

CATIA 3D 参数化变量造型器使设计师可以对 CATIA 几何增加参数数据,并捕捉到设计意图,加快产品开发。这项技术使设计更改的处理自动化,它使得在维持原有设计意图和约束不变的前提下,在快速开发设计替代时具有更大的灵活性。3D 参数化变量造型器能使设计师在整个设计上或是在部分设计上增加参数时作出选择,设计师也可以灵活确定什么时候增加这些参数。

#### (9) 特征设计(Feature-Based Design)

CATIA 特征设计产品通过把系统本身提供的或客户自行开发的特征用同一个专用对话框结合起来,从而增强了设计师建立棱柱件的能力,这个专用对话框着重于一个类似于建立一族可重新使用的零件或用于制造的一个设计这样一个单一过程。

#### (10) 装配设计(Assembly Modeling)

CATIA 装配设计产品可以使设计师建立并管理基于 3D 零件和约束的机械装配件。装配件可以由多个主动或被动模型中的零件组成。零件间的接触自动地对连接进行定义,方便了 CATIA 运动机构产品进行早期分析。基于先前定义零件的辅助零件定义和依据其之间接触进行自动放置,大大加速了装配件的设计,后续应用可利用此模型进行进一步的设计、分析、制造等。

#### (11) 复合材料铺层设计(Composites Covering)

CATIA 复合材料铺层设计产品和制造工程师在零件模型上定义、管理和分析复合材料相关的数据。通过信息的相关性,工程师可以在零件几何上增加技术信息,作为一个单独的元素构造出复合材料的铺层。该产品同时支持复合材料的设计、生产计划和工艺规程。所提供的分析和可视化工具可改善所生成的零件的质量。系统可自动生成生产加工过程中所需的如铺层顺序、重量和面积、铺层下料边界、材料厚度等信息的报表。

#### (12) 创成式复合材料铺层设计(Generative composites Covering)

创成式复合材料铺层设计可使复合材料设计师根据结构分析要求定义复合材料零件。在复合材料零件上完成的初始有限元分析的结果就是每个元素的分层特性。在同一铺层结构零件上确定所有的元素,即可定义复合材料零件的不同区域。区域定义以后,创成式复合材料铺层设计产品即可应用企业对叠层顺序和铺层定义的设计规则,自动生成铺层几何,并在每一层上附加属性信息。当铺层生成以后,可使用已有的 CATIA

的复合材料铺层设计产品进行详细零件设计,进一步细化铺层的几何。直接使用集成在复合材料铺层设计产品用户界面上的创成式复合材料铺层设计(GCC)的功能,可使工作效率更高、更容易使用。该工具的目的是针对在设计早期阶段要进行许多设计变化的情况,自动进行复合材料零件的定义。创成式复合材料铺层设计(GCC)可减少目前手工处理的错误,快速推出高质量的初始设计。

#### (13) 创成式复合材料铺带设计(Generative Composites Tape Laying)

创成式复合材料铺带设计模块可使工程师定义每个铺层由几片单一方向铺带材料构成的复合材料,铺带作为铺层特征来定义,利用技术参数自定义铺带几何,基于铺层轮廓和一组窜差,相邻铺带之间空的重叠以及每层内部的纤维方向自动优化,与其他复合材料应用完整集成,可为用户提供真实纤维路径分析,加工过程和激光投影系统编程的支持。

#### (14) 钣金设计(Sheetmetal Design)

CATIA 钣金设计产品使设计和制造工程师可以定义、管理并分析基于实体的钣金件。采用工艺和参数化属性,设计师可以对几何元素增加象材料属性这样的智能,以获取设计意图并对后续应用提供必要的信息。

#### (15) 创成式航空钣金设计(Generative Aerospace Sheetmetal Design)

创成式航空钣金设计产品满足航空航天行业钣金零件设计师的需求,它提高了成形件及展开件的设计效率,也改善了设计质量,它也可以捕捉公司的设计知识,包括设计和制造约束,方便了液压成形结构件的设计。该产品也提供了一个专用于航空钣金零件的预定义特征编目。

#### (16) 3D 实用尺寸和公差标注(3D functional Dimensioning and Tolerancing)

CATIA 3D 实用尺寸和公差标注方便了精确实体及近似实体模型的公差指定。该产品可自动地建议在与装配实际说明相一致的地方标注公差,并提出与所选择曲面相吻合的公差类型。3D 实用尺寸和公差标注产品对整个公差标注的质量提高大有帮助,并通过避免过于严格的公差要求,有助于降低生产成本。同时该产品支持有关公差规则和语法的国际标准。

### 4.2.2 外形设计和风格设计

#### (1) 曲面设计(Surface Design)

CATIA 曲面设计产品使设计师能够快速方便地建立并修改曲面几何。它也可作为曲面、面、表皮和闭合体建立和处理的基础,曲面设计产品有许多自动化

功能,包括分析工具,加速了曲面设计过程。

#### (2) 高级曲面设计(Advanced Surface Design)

CATIA 高级曲面设计产品提供了可便于用户建立、修改和光顺零件设计所需曲面的一套工具。高级曲面设计产品的强项在于其生成几何的精确度和其处理理想外形而无需关心其复杂度的能力。无论是出于美观的原因还是技术原因,曲面的质量都是很重要的。设计师都能从自动倒角和曲面历史修改这些特有功能中收益非浅。

#### (3) 自由外形设计(Free Form Design)

CATIA 自由外形设计产品提供设计师一系列工具,来实施风格或外形定义或高度复杂的曲线和曲面定义。对 NURBS 的支持使得曲面的建立和修形以及与其他 CAD 系统的数据交换更加轻而易举。实时分析工具加快了设计过程,并可生成更高质量的设计。

#### (4) 整体外形修形(Global Shape Deformation)

CATIA 整体外形修形提供了一套工具,使用户在 CATIA 模型中使用表皮、面、曲面和曲线对复杂的外形进行连续的修形。用户在工具编目中选择合适的工具,以此对 CATIA 元素进行操作。这些整体和非线性修形(比如拉伸、弯曲和扭曲等)仍使 CATIA 元素保持特征线和几何上的连续性。用户控制包括位置公差、曲面细分和修改程度等这样一些关键性“承前启后”参数。

#### (5) 创成式外形建模(Generative Shape Modeling)

创成式外形建模产品是一个进行曲面设计的先进工具。该产品通过对设计方法和技术规范的捕捉和重复使用,可大大加速设计过程。在曲面技术规范编辑器中捕捉到的设计意图,可使用户在设计周期中的任何时候,方便快速地实施重大的设计更改。产品制造、分析和工程图纸的生成等后续的应用都可与曲面的定义过程中的更改相关联,从而极大地提高了效率。创成式外形建模产品可广泛应用于汽车白车身设计、飞机外形设计以及日用消费品塑料件的设计中。

#### (6) 白车身设计(Body-in-White Templates)

白车身设计产品提供了一种新的设计方法,可用于设计汽车内型板和车体加强肋等复杂的薄板零件。白车身设计提供有高效的,并支持设计团队协同设计和重复使用设计和制造技术规范的工作方式。通过 3D 曲线等形状的扫掠,可自动生成曲面,并具有很高的曲面质量,避免了耗时的重复设计。该新产品使 CATIA 系统已有的混合造型技术更加完善。

#### (7) 曲面展开(Develop)

CATIA 曲面展开产品可使设计工程师快速地将计算钣金工的 3D CATIA 模型的相应展开形状,相反也可以从已经修改的展开形状重新建立起 3D 形状。

通过两种可选的变形分布功能,设计师对各向同性材料和各向异性材料都能够进行处理。用户通过交互式地修改展开的平面图形或 3D 模型,可以分析展开形状的变形和制造工艺性。

#### (8) 测量设计(Geodesic)

CATIA 测量设计产品使设计师能够遵循定义的规则在曲面、面和表皮上直接绘制曲线。用户可以在复杂形状的钣金或防护罩上设计腐蚀区,同时也可设计复合材料零件上的铺敷。该产品也可使用户方便地计算出复杂曲面、面或表皮上的距离。

#### (9) 刻模/压模(Engraving/Embossing)

刻模/压模产品可使用户从 2D 位映像图快速地生成 3D 刻模或压模曲面。也可以将 3D 徽标或参考号冲压或刻压在机加零件上。所提供的喷涂盒方便了用户在零件曲面上将其图像转换为复杂 3D 设计前,用标准的绘图工具创造性地草绘。该产品可使设计师快速地编辑现有光栅数据,直到像素级。该产品可对许多不同类型元素包括表皮、面、曲线和点数据云团等进行操作。设计师可以选择一些容易理解的选项来控制刻模/压模的深度或高度、雕刻曲面的凸度及开模方向等。而其中所包括的一些实用程序可使设计师在刻模或压模操作之前能够对 2D 图像进行建立、选择、验证和编辑等。

#### (10) 由点数据云团到几何模型(Cloud to Geometry)

CATM 由点数据云团到几何模型产品提供了由物理世界和 CATIA 数字化样机之间的桥梁。该产品可使设计师将物理样机转换到 CATIA Designs 下并转变为数字样机,并将测量设备数据转换为 CATIA 数据。该产品同时提供了一套有价值的工具来管理大量的点数据,以便进行过滤、采样、偏移、特征线提取、剖截面和体外点的剔除等。由点数据云团到几何模型支持由 CATIA 曲线和曲面生成点数据云团。反过来,也可由点数据云团到 CATIA 曲线和曲面。

#### (11) CATIA—ALIAS 数据互操作接口(CATIA—ALIAS Interoperability)

对于外形至关重要的行业,比如汽车和日用消费品, CATIA—ALIAS 数据互操作接口这个新产品可在 CATIA 和 Wavefront 的 ALIAS 间提供有效的高质量数据交换,它提高了风格造型过程的效率,同时保证了这些行业的设计师与工程师间更方便的协调设计。该解决方案避免了会导致耗时的模型清理的数据传输错误,行业设计师和工程师可以更高效地交换设计概念,提高了设计质量,并在更短的时间内完成项目。

### 4.2.3 分析与模拟

#### (1) 创成式零件应力分析(Generative Part Stress Analysis)

CATIA 创成式零件应力分析产品在产品开发过程初期提供设计师一个透明的应力分析工具,作为铸件、锻件或厚壁零件设计的指导。交互式地对零件而非有限元模型进行操作,设计师只用 6 个简单的步骤在零件上进行应力检查。

#### (2) 创成式零件优化(Generative Part Optimization)

CATIA 创成式零件优化产品可在更少的时间内生成优化产品,用户只需输入参数,系统可对设计自动优化,载荷及约束值随着每次迭代,自动地显示出来,用户可通过历史浏览器对所有分析结果的变化进行研究,若同 CATIA 创成式零件应力分析产品共同使用,创成式零件优化产品可使用户边设计边分析,同时可立即获得有关质量、位移以及主应力等数据,即使不是一个有限元分析专家,也可使用该产品。

#### (3) 创成式零件动态分析(Generative Part Dynamic Analysis)

CATIA 创成式零件动态分析产品可使用在零件上进行动态(模态)计算,从而可对由振动引起的关键问题进行深入研究,由共振零件模态频率引起的问题可在开发过程初期发现,从而避免了昂贵的重新设计。由于其便学易用,设计师、工程师或分析师都可使用 GPD,只要简单地选择一个图标,用户就能进行模拟输出,以便对分析结果能更好地理解。CATIA 创成式分析产品早已支持静态分析及设计优化,GPD 进一步扩展了这些分析产品的使用范围,GPD 也同样采用了用于整套 CATIA 有限元分析产品中的有效网格划分及求解技术,但它对用户却是透明的。

#### (4) 分析基础框架(Analysis Infrastructure)

CATIA 分析基础框架为其他独立的 CATIA 分析产品提供了基础功能。此外,它集成到 CATIA 环境下的其他任何希望的求解器。公司可选用适当的求解器进行机械、热力学、声学、动力学或其他方面的分析计算:

- 集成化的 CATIA ELFINI 求解器
- MSC/NASTRAN
- ANSYS
- 其他商品化求解器
- 公司专用求解器

#### (5) 有限元模型生成器(Finite Element Modeler)

CATIA 有限元模型生成器产品作为一个完整的工具,可为诸如机械和热力学这样的许多方面的分析

准备几何模型。该产品同时具有强大的网格划分功能,配以自动化的特征,便可生成有限元模型。有限元模型生成器具有开放式体系结构,可以同其他商品化或专用求解器进行接口。该产品同 CATIA 紧密地集成在一起,简化了 CATIA 客户的培训,并大大有助于在一个强大的 CAD/CAM/CAE 系统中完成整个有限元模型造型和分析的解决。

#### (6)ELFINI 解算器(ELFINI Solver)

CATIA ELFINI 解算器产品是一个专用于求解线弹性和热问题的有限元解算器。这个产品能进行复杂的线性、动力、屈曲或静力分析,并能对接触进行模拟。CATIA ELFINI 解算器支持各种材料类型和载荷,另外提供了一个强大并可用的有限元元素库和一系列输出选择。

#### (7)科学表示管理器(Scientific Presentation Manager)

CATIA 科学表示管理器是用于数据可视化的一个集成化的工具。该产品主要用于有限元求解器如 CATIA ELFINI 求解器产品等所进行的计算结果的后置处理。科学表示管理器广泛的几何功能使其可以给出基于网格的数据。用户可以用许多方法包括对结果的颜色标码、标注、重叠显示和模拟等均可快速、直观地处理和显示这些生成数据。

#### (8)ANSYS 接口(ANSYS Interface)

CATIA ANSYS 接口产品能对 CATIA 有限元模型生成器所建立的模型进行预处理,以供 ANSYS 求解器使用。ANSYS 接口产品将 ANSYS 通用求解器集成到 CATIA 环境中,包括对 ANSYS 求解器生成的数据进行后处理,以使用 CATIA 科学表示管理器产品进行描述。已熟悉 ANSYS 产品的用户和已熟悉 CATIA 产品的用户都将从这个结合中受益匪浅。

#### (9)NASTRAH 接口(NASTRAN Interface)

CATIA NASTRAN 接口产品能对 CATIA 有限元模型生成器的数据进行预处理,以供 MSC/NASTRAN 求解器使用。另外这个产品还可对 MSC/NASTRAN 的分析数据进行后处理,并通过 CATIA 科学表示管理器进行可视化显示,NASTRAN 接口将 MSC/NASTRAN 求解器集成到 CATIA 中,使处理过程更加顺畅,增强了数据的集成性,并减少了额外培训的需求。

#### (10)运动机构(Kinematics)

CATIA 运动机构产品可使用户通过真实化仿真设计并验证机构的运动状况,基于新的或现有的零件几何,运动机构可以建立许多 2D 和 3D 连接。设计师可以模拟机构运动,来分析加速度、干涉、速度、间隙等等。

#### (11)装配件与部件的结构分析(Assembly and Substructure Analysis)

CATIA 装配件与子结构分析产品是专为必须对典型装配件组成的大结构进行分析的机械工程公司设计的,为了解决这些问题,装配件与子结构分析为定义不同零件的结构耦合提供了管理工具。该产品可在大型系统上实施并行工程,也可根据装配中的上下相关关系对部件进行快速设计分析迭代。

### 4.2.4 机械制造

#### (1)制造基础框架(Manufacturing Infrastructure)

CATIA 制造基础框架产品是 CATIA 数控产品的基础。该产品提供有一系列工具,可用来在电子样机上模拟重放和验证加工走刀路线。其他工具还可用来采用图形化的方式检查和修改人工不可读格式的加工走刀轨迹。用户可定义和管理用于操作的 CATIA 宏指令,以及定义和管理后置处理的词汇和语法。

#### (2)工具库存取(Toolshop Access)

CATIA 刀具库存取对用户提供一个实时环境,以运行和管理将 CATIA 刀具轨迹转换成机器专用的 NC 代码文件所需的各种作业。用户可以编辑、拷贝、更名、删除和存贮自己的 NC 文件。该程序可使用户检查批处理作业执行报告,改变作业执行的优先级,从作业队列中删除不需要的作业等。刀具库存取提供易用的 Motif 用户界面,同时也提供有后置处理器样本,使用户可检验或改善其 NC 方案。

#### (3)后置处理构造器(Post-processor Builder)

CATIA 后置处理构造器产品可使用户建立自己的数控(NC)后置处理器。该程序的使用并不需要专门的编程知识,同时也附加了后置处理器样本以供调试用。后置处理构造器同其他的制造产品互为补充,并支持多轴及定轴 NC 机床,该程序可进行一致性检查以及 NC 批处理和文件管理。

#### (4)多轴铣(Multiple Axis milling)

CATIA 定轴铣产品可使数控程序员生成、验证、实施或维护铣床和多工序数控机床的走刀路线。通过走刀路线编程步骤的自动化以及模拟各种简单或复杂 2.5 轴和 3 轴(包括有旋转台面的 3 轴)机床的铣或钻操作。定轴铣产品简化了机床编程过程。

#### (5)粗加工(Rough Cutting)

对定轴铣和多轴铣产品而言,CATIA 粗加工产品是一个集成化的选项。正因为系统与其他 NC 铣产品的集成,用户只需集中精力于加工目的,NC 编程员可以对成形件或粗毛坯材料采用许多平面粗加工技术,同时可以完全防止碰撞。由于减少了交互式操作的次数,用户生成的零件可以为精加工留出最小量的材

料。

#### (6) 铣加工分析(milling Analysis)

CATIA 铣加工分析产品使数控(NC)编程员可以优化刀具选择和零件定位,以提高铣加工的效率,并在实际加工之前可确认先前计算出的走刀路径。该产品由两个分析工具组成:一是预切分析,用来估算不同刀具选项的效果。另一个是切后分析,用以确定需要清除的区域。对于使用3到5轴铣床的NC编程员,在它们加工锻模件、压模件和注模件以及其他一些具有不便接近区域的零件时,可以提高加工效率,降低费用。

#### (7) 注模和注模加工辅助器(Mold and Die Machining Assistant)

CATIA 注模和压模加工辅助器可注模和压模类的零件自动定义数控(NC)程序。用户可利用已预先定义好的加工方法,在零件上附加特定的几何信息。由于使用经过验证的加工方式,这种方式简化了程序员的工作,并可提高精加工的质量。在处理过程的最后,系统自动生成NC技术文档。

#### (8) 2轴半加工编程器(Prismatic Machining Programmer)

CATIA 新的2轴半数控加工编程器产品对CATIA已有2轴半数控加工辅助器产品的补充,该产品提供了针对简单零件的钻孔和铣削等加工操作的NC编程功能。根据所选择的几何元素,用户通过查询存放有企业专有的加工环境的工艺数据库,生成加工操作。这样,就定义了加工操作的所有准确信息,包括机床和刀具、加工参数(进给速度和旋转角度)以及加工类型(如钻孔、平面铣削和点位加工等。)在加工操作定义和修改完成以后,用户可以显示刀具走刀轨迹,以便检查加工参数。

#### (9) 曲面加工编程器(Surface Machining Programmer)

CATIA 曲面加工编程器产品可帮助用户进行3坐标铣削加工操作的数控编程。该产品将久经验证的CATIA数控铣削加工技术和新的CATIA制造产品平台结合起来。这就使编程员可以使用工艺数据库和标准化的加工操作,以及CATIA综合加工环境中的其他所有信息。这个环境综合了零件几何的信息、毛坯、夹具、机床和设置参数等信息,提供了企业加工操作的全面描述。

#### (10) 多坐标加工数控编程器(Multi-Axis Machining Programmer)

CATIA 多坐标加工数控编程器产品在CATIA的制造产品系列中采用了新的多坐标数控编程功能,并应用NCCS(数控计算机科学)中最先进的技术用于5坐标的加工计算。该产品是CATIA 2轴半加工编程

器和CATIA曲面加工编程器的强大补充。这些产品共同组成一个可满足从2.5轴到5轴铣削加工和钻孔加工的复杂零件数控加工需要的完整解决方案。

#### (11) 2轴半加工辅助器(Prismatic Machining Assistant)

CATIA 2轴半加工辅助器产品对基于精确实体的2轴半零件的辅助编程提供了强大的工具。无论是从零件准备期的加工特征辨认,还是到制造数据的输出和确认,该产品实际上涵盖了整个设计/制造过程。由于CATIA NC工艺数据库和过程模型的支持,2轴半加工辅助器采用专用的铣、钻技术,简化了NC编程员的工作,降低了零件制造成本,改善了零件质量。

#### (12) 车加工(Lathe)

CATIA 车加工产品是数控车加工编程工具,它采用的是曲线架、曲面或实体模型而来的2D或3D轮廓图。系统将整个编程过程划分为一系列简单的任务,每个任务作为一个简单的操作如轴向粗加工或靠模铣精加工。

#### (13) 车加工编程器(Lathe machining Programmer)

CATIA 车加工数控编程产品专用于数控车削加工,可应用于一般机械、航空航天和汽车行业。该产品可使用户在MFGPROG功能中定义并管理新一类的针对车削加工的机械加工操作。这些操作在零件操作(Part Operation)功能中生成的,可以包含有基于制造基础框架产品的所有制造产品中机械加工操作。

#### (14) 规划检测辅助器(Inspection Planning Assistant)

CATIA 规划检测辅助器(IPA)是一个基于特征的引用产品,可用于方便地定义坐标测量机(CMM)可读的检测计划,控制所工完成的零件在坐标测量机上检测路线。在高度自动化的3D FT&D分析的基础上,可根据零件上已经定义的3D公差,建立一系列控制特征,对于没有预先定义3D公差的零件,可在所选择的曲面或实体几何元素上生成一个探测特征。在控制特征和探测特征定义以后,就可生成一个坐标测量机可读的DMIS格式的启动程序。要完成一个完整的CMM程序检测路径,需考虑CMM测量机类型和装配因素。这项工作可由CMM软件完成,也可由CATIA规划检测辅助器提供的脱机编程功能来完成。

#### (15) STL快速样机(STL Rapid Prototyping)

STL快速样机是一个专用于通过STL过程生成快速样机的CATIA产品,该产品对专家或非专家均是有用的,它可将CATIA几何图形转换为标准的STL文件格式。而生成的高质量文件可产生高度精确的样机,该产品可将一个大型的样机设计划分为更小



的部分,这样,经过几个步骤便可快速建立起样机。对于注模样机,偏置功能可在一个真实零件上增加一个厚度。用户可以通过手工修改 STL 文件,甚至从其他系统上也是可以的

(16)工作单元设计和机器人编程(Cell Design and Robot Programming)

CATIA 工作单元设计和机器人编程产品通过有效的工作单元设计及脱机机器人及控制器的编程及仿真,可以方便快速地建立从事一个给定工作的机器人。该产品提供了许多方法以验证在一定工作条件下机器人的适应性,然后生成所需的程序。

(17)机器人和控制器定义(Robot and Controller Definition)

CATIA 机器人和控制器定义产品用来定义机器人和控制器特性及行为机理。该产品提供有开放式体系工具,可使机器人和控制器的制造商或集成商增加专利技术和专用算法。以取得高度准确的行为模拟。

(18)钣金排料(Nesting)

CATIA 排料产品对于在板材上优化一组零件的自动或手工定义提供交互式功能。该产品通过减小原材料的消耗,提高产品设计师和制造工程师的效率,从而降低了制造成本。

#### 4.2.5 设备与系统工程

(1)空间分析

CATIA 空间分析产品允许汽车、航空航天、造船和设备行业的机械设计师检查零件干涉和验证 CATIA 3D 元素之间的间隙。该产品对快速方便地在混合环境下辨别和确定相关干涉提供了集成化的工具,从而增强了专业间的协作。

(2)空间漫游

CATIA 空间漫游产品可模拟化引导用户围绕或进入他们的 3D 设计中行走。空间漫游可使设计师更加高效,它仅显示由设计师定义的工作体内的那些模型零件,而设计师只需对这些零件进行操作即可。

(3)系统路径定义

使用 CATIA 系统路径定义产品,用户可以在 CATIA 数字化样机中为象管路或电线束这样的元素定义一个 3D 网络。系统可自动寻找最佳的网络路径,并自动地检查每个结点的连接性,同时把路径与支持约束联系起来考虑。系统也可对用户定义的规则进行检查,以保证符合技术要求。该产品开放式体系结构可使公司集成本企业的技术经验,并对产品进行广泛的客户化,作为集成化解决方案的一部分,用系统路径定义所建立的路径随后可以被 CATIA 电线束安装产品所使用,以此可自动定义电线束路径。

(4)系统原理图

CATIA 系统原理图产品可使用户快速建立和修改任何类型的原理图,包括电子原理图、布线图、液压传动图和处理过程图等。系统原理图满足设计智能原理图布局或 2D 路径定义的工程需要,例如简图、基本原理图和 DI 艺流程图。

(5)知识工程

CATIA 知识工程产品可使用户通过智能化规则定义和检查功能,将公司技术经验深入他们的设计过程。设计师和绘图员在管理复杂的零件装配时,也清楚这样是符合技术规程、安全条例和公司标准的。

(6)装配模拟

CATIA 装配模拟产品可使用户定义零件装配或拆卸过程中的轨迹。使用动态模拟,系统可以确定并显示碰撞及是否超出最小的间隙,用户可以重放零件运动轨迹,以确认设计更改的效果。

(7)电气设备和支架造型

CATIA 电气设备和支架造型产品为设计师提供了建立电气标准件库的工具。以此产品建立的库专用于 CATIA 电线束安装产品。

(8)电线束安装

CATIA 电线束安装产品允许用户在大型的或复杂的装配件中方便地设计和铺设 3D 线束。该产品的若干特点可使铺设电线束的设计、连接、修改和分析自动化。在同 CATIA 电气设备和支架造型产品所建立的设备库一起优化使用时,电线束安装可使用户处理电线束的参数和约束,而系统对几何图形进行管理。

(9)电气生成模板

CATIA 电气生成模板产品使用户可生成 2D 模板设计以用于电线束的制造,它是 CATIA 电线束安装及 CATIA 电子设备与支架造型的姐妹产品。使用以电线束安装产品建立的 3D 设计也可以同时使用以电气设备和支架造型产品所建立的库中的电气设备和支架零件,电气生成模板产品可计算并分析线束的 2D 模板布局。

(10)3D 管道设计

CATIA 3D 管道设计产品可以帮助管道设计师铺设、修改并分析管路和管道网络,汽车、航空航天、造船、设备、仪表和工厂设计行业的设计师可以迅速、轻松地开发出集成于装配体中管道或塑料管路的复杂网络。

(11)船舶设计

CATIA 船舶设计产品使设计师能快速建立、修改和处理在船体设计中代表钢加强板的复杂 3D 实体参数化零件。借助一个专用的交互式用户界面,设计师可以增加元素以建立用户定义的船体断面。在用户提出



修改后,系统对智能化的 3D 参数化对象进行操作,以更新相关零件,并生成准确的生产数据报告。

#### (12)船舶生产

CATIA 船舶生产产品抽取与 3D 船体模型相关的数据以生成装配图和详细设计图、零件轮廓图和排料参考图。系统对这些生产中关键要素的生成确保了模型设计的准确性,从而也就改善了生产质量。将船舶生产与类似于船舶设计、排料、制图、库和机器人模块等其它 CATIA 产品结合起来使用,可最大限度地让利于用户。

#### (13)2D 建筑设计

CATIA 2D 建筑设计对建立、修改和管理 2D 建筑部件提供了一套综合性交互式工具。设计师可以对复杂的建筑项目准确地建立起详细的楼层布置图或竖直剖视图。该产品对设施布局的更改比如办公室装修及工业综合结构等的管理也是很有帮助的。由于用户是同定义有工艺属性的智能化部件打交道,而各部件间的相关性使建筑师和工程师可以轻松的管理更新。借助于 CATIA 3D 结构件产品,用户可以方便地从 2D 平面图转化到 3D 投影图。

#### (14)3D 结构件

CATIA 3D 结构件可使设计工程师构造在设备、储存架、其它器械和建筑中常见的复杂的框。使用构件建立工具,用户可以定义常见部件以及设计、修改并分析由标准件和常见构件组成的结构件。由于这些产品完全集成,用户可容易地在 2D 平面图和 3D 投影图之间进行转换。

### 4.2.6 CATIA 网络计算

#### (1)CATweb 空间漫游器(CATweb Navigator)

CATweb 空间漫游器产品提供了通过任何基于 TCP/IP 协议的网络访问 CATIA 数据的能力,即使在低带宽调制解调器下也可快速显示大型数据,可通过本地控制来获取在最合适的视窗内显示二维和三维数据的能力。该产品既可浏览现有的 CATIA 输出文件,也可远程生成 CATIA 输出文件,新的产品还可访问 VRML 格式数据和 AEC 输出结果。

#### (2)CATweb 空间分析器(CATweb Space)

在 Web 漫游器上进行 CAD 的剖切面检查和测量计算。CATweb 空间分析(CATweb space)做为 CATweb 漫游器的附加产品,可允许审查人员(包括非 CAD 用户)在复杂的 CAD 模型和装配件上进行测量分析。CATweb 空间分析可在 CAD 模型和装配件上进行操作和准确地定位剖切面显示,以便进行审查位置、大小、间隙和干涉等高级分析。可使用很高的精度,检测 3D 元素的干涉情况。具有 CATweb 漫游器的

优点:可使用常用的标准的商用计算机和 workstation 访问 CAD 和 PDM 数据。使用 Internet、Intranet、LANs WANs 等 TCP/IP 网络环境。

#### (3)CATweb 信息编辑器(CATweb Publish)

专家型的 HTML、CATweb (CATweb Publish)做为 CATweb 漫游器的附加产品,可以产生含有 CATIA 或 ENOVIA 数据的网页。不必是管理层用户或 Webmaster 即可出版含有 CATIA 图片、3D 模型、或 PDM 元数据的 HTML 网页。通过抓取 CATIA 模型视图并插入到准备好的 HTML 页中,大大简化了 CATweb 的处理过程。

### 4.2.7 应用体系

#### (1)CATIA 目标管理器(CATIA Object Manager)

CATIA 目标管理器产品包含所有 CATIA 解决方案产品和配置所必需的完整的基础框架和共同特点。该产品提供了共享用户界面、数据管理和环境控制特点,这些对一个完整的、CAD/CAM/CAE 系统的有效操作是必需的。

#### (2)CATIA 浏览服务(CATIA Viewing Services)

CATIA 浏览服务产品对一些特定的 CATIA 产品提供了低成本的基础,在这些 CATIA 产品中,基于工作站的报告建立、浏览和标注是主要的需求。CATIA 浏览服务是为设计支持人员定制的,他们并不一定会建立、分析或修改几何图形。这样的支持功能可能包括采购、市场和文档部门,在这里处理的是审查、标注和技术以及桌面排版等事务。

#### (3)库管理(Library)

CATIA 库文件产品存放、检索和管理几何部件,以备在 CATIA 模型中重复使用。部件制造商可能希望与其客户共享他们库的拷贝。此库的存在促使客户在总装时使用制造商的部件,有了库产品,用户可以避免重新设计现在的 CATIA 对象,由此提高了生产率,减少了错误。

#### (4)远程协作(Conferencing Groupware)

CATIA 协商组件使在物理世界中相距遥远的人可以一起工作来审阅 CATIA 设计。该产品提供了开始、邀请、参加和退出一个协商的管理工具。闲聊工具可通过文本、音像来进行通讯。CATIA 图像可捕捉共享。2D 白板可使用户同时浏览并标注 2D 图像,就像浏览 CATIA 模型那样。当同 CATIA 4D 漫游器产品一起使用时,用户可同时对 CATIA 3D 模型进行浏览、标注和处理。

#### (5)CATIA 分布式服务(CATIA Distributed Services)

CATIA 分布式服务对 IBM RISC System/6000 工程工作站和 MVS 或 VM 主机系统之间提供关键的互操作功能。CATIA 分布式服务使用户可在一个工作站上执行主机上的 CATIA。该产品可使用户进行本地或远程数据传输或转换,也可不经数据传输而直接对主机文件进行操作。要进行协同工作, CATIA 网络用户可在浏览一个模型的同时,通过讨论窗口进行评论。用户如果在工作站上有一个 CATIA 数据管理会话,而同时将 DB2 数据库在 MVS 主机系统上用作中央数据仓库,就可以对 DB2 数据库中的数据进行管理。

(6) CATIA-CADAM 2D 标记和标注 (CATIA-CADAM 2D Markup and Annotation)

CATIA-CADAM 2D 标记与注释是一个独立产品,它可使用浏览或交互式“图阅”图样,该产品价格低廉,使用方便,应用于并行工程或无纸制造环境中。用户可以浏览和标注在 CATIA-CADAM 制图中建立或输入到 CATIA-CADAM 制图中的矢量 (CAD)、光栅 (图样) 或混合图样。

(7) CATIA 混合光栅 (CATIA-CADAM Hybrid Raster)

CATIA 混合光栅产品可使用户在 CATIA 模型中输入、建立并编辑光栅数据。系统可以保存包含光栅和矢量两种几何的模型。对于多资源用户,混合光栅编辑器可在与 CATIA 系统集成的单一数据库中进行文件的检索与归档。CATIA-CADAM 混合光栅产品可以归档、更新并管理由图纸生成的光栅图像,用户可通过光栅编辑和(或)光栅图像转换完全地或有选择地将光栅图像更改为矢量几何,然后用户就可用 CADAM 制图产品对图像进行编辑,CADAM 制图产品可以恢复图样的准确性,在使用 CADAM 制图变量式设计功能时,CATIA-CADAM 混合光栅是从图纸到设计模型的唯一链条。

(8) 实体样机 (Solid Mock-up)

CATIA 实体样机产品允许在设计过程的早期快速方便地建立实体样机,以完全支持零件集成在数字化样机。样机实体在性能和效率方面优于精确实体,使其成为大型装配样机中最合适的几何表达。再者,象在装配研究和概念可视化这种情况下,并不需要精确曲面边界定义。对于预设计,在样机实体上进行大量分析是足够准确的。

(9) 可视化工作室 (Visualization Studio)

CATIA 可视化工作室可由 CATIA 3D 模型生成具有非凡质量的照片般真实的图像。采用强大的光线追踪技术,系统可以对光的行为生成自然正确的仿真。通过在网络和多处理器机器上实施并行算法,可视化工作室可提供快速的响应,通过真实化渲染,设计师可

显示复杂的模型,应用市场销售及培训方面。建筑师可展示新建筑设计和内部装修概念的效果。采用具有高度准确的反射及折射展览室旋转合效果,设计师可以检查曲面质量及设计美观程度。

(10) 真实化渲染 (Realistic Rendering)

通过系统提供的渲染参数集, CATIA 真实化渲染可快速方便地生成 CATIA 模型照片般真实的图像。这些参数包括照相机、光源、渲染风格、电影和展览室等。用 CATIA 可视化制作室产品可以开发新的渲染参数集,通过高级特性渲染和抽样算法,真实化渲染可生成同 VST 产品同样卓越的性能。通过对已知环境反射的估算,设计师可进行曲面分析。

(11) 图像设计 (Image Design)

CATIA 图像设计是 CATIA 可视化制作室的替代产品,在不需要光线追踪和其他高级技术时,它可以帮助设计师、技术资料编写人员以及市场部门的人员建立起 CATIA 模型的图像。用户可以从所提供的库中选择 3D 体素,也可以开发自己的体素,然后将这些体素安排到模型上。光源、阴影、颜色及其他技术增强了图像的真实感。图像设计也可提供静态曲面分析(对应于 CATIA 自由外形设计产品所提供的动态工具)。

(12) 4D 漫游器 (4D Navigator)

CATIA 4D 漫游器可使用户在非常大型的机械和建筑 CAD 模型中进行可视化漫游。它允许在装配的早期检查时,对设计进行交互式 and 动画可视化检测。4D 漫游器提供方便使用的直观控制,以简化漫游和参数调整。用户在开始图形化检查时无需进行任何初始操作,便可直接装载 CATIA 模型、CATIA 会话或 CATIA 数据管理 Access 装配。而客户机/服务器体系结构和智能化数据仓储机制可带来非凡的速度。

(13) CATIA-CADAM 耦合器 (CATIA-CADAM Coupler)

CATIA-CADAM 耦合器产品在 CATIA 和 CADAM (主机 CADAM 或 PROFESSIONAL CADAM) 环境间对数据传输提供了易用、高度可靠的双向桥梁。

(14) DXF/DWG 接口 (DXF/DWG Interface)

CATIA DXF/DWG 转换器可使用户在 Micro CADA MX/6000 (日本版本可到 Version 2) 和 CATIA-CADAM 制图之间进行模型的输入或输出。

(15) CATIA-CADA MCX 调转换器 (CATIA-CADAM MCX Translator)

CATIA-CADAM MCX 转换器可使用户在 Micro CADAM X/6000 (日本版本可到 Version 2) 模型输入到 CATIA-CADAM 制图产品中。

(16) 信息编辑器 (Publishing)

CATIA 信息编辑器产品可使用户以符合行业标准的出版格式生成 CATIA 输出。用户可以生成材料清单、零件列表和产品文档等报告。信息编辑器产品可以生成分解装配显示并捕捉 CATIA 会话屏幕以在产品文档中使用。用户也可以将信息编辑器产品作为一个转换界面,将一种标准格式直接转换成另一种标准格式。

#### (17)STEP AP203 接口(STEP AP203 Interface)

无论是何种硬件系统和软件系统,CATIA STEP AP203 接口可使用户在异构环境中交换 CAD/CAM 模型数据。基于新兴的产品模型数据交换的国际标准(STEP),用户可以获得更好的精确度、可靠性和一致性。在用 CATIA 模型建立一个 STEP 文件或从 STEP 文件生成一个 CATIA 模型时,STEP AP203 接口采用新兴的 ISO 标准中性文件格式,它打开了不同标准和专用方法带来的屏障。公司及其代理商也可以相关信息流进行实时数据交换。

#### (18)STEP AP214 基础框架接口(STEP AP214 Infrastructure Interface)

STEP AP214 基础框架接口将 CATIA 模型数据转换成 AP214(应用协议)STEP(产品模型数据交换标准)文件格式,同样可将 AP214 STEP 文件格式数据转换成 CATIA 模型数据,以实现异构 CAD/CAM 系统间先进、可靠并独立地数据交换。AP214 主要是为汽车行业开发的,并且大大方便了承包商与其子承包商间的数据交换。AP214 协议同时考虑到了实体和各种各样的曲面,STEP 确定要贯穿产品的整个生命周期。

#### (19)IGES 集成化接口(IGES integrated Interface)

CATIA IGES 集成接口通过中性 IGES 格式同其他 CAD/CAM/CAE 系统双向交换数据,支持行业标准的 IGES 格式(目前是 5.2 版本)减少了数据传输和清理时间,这就提供了完整的 IGES 功能,无需另外 CATIA IGES 支持产品,这一单一 CATIA IGES 方案减少了费用,简化了安装及管理资源。

#### (20)IGES 接口(IGES Interface)

CATIA IGES 接口产品采用初始图形交换规范(IGES)中性格式在 CATIA 系统和其他 CAD/CAM/CAE 系统之间进行数据转换。

#### (21)IGES MIL 标准 28000 接口(IGES MIL Std 28000 Interface)

CATIA IGES MIL Std 28000 接口产品可使用户在产品的开发及整个生命周期中转换军用 D28000A IGES 中性格式文件生成 CATIA 模型。只要某些模型元素保持完整,用户可以由新生成的 CALS 28000A

模型重新转换为 CALS MIL D 28000A IGES 中性格式。这一双向转换确保用户能够满足所要求的政府通信标准。

#### (22)图形交互式接口(Graphics Interactive Interface)

CATIA 几何交互式接口(GII)产品对开发 CATIA 交互式功能提供了应用编程界面和高级语言。该产品提供了一致性的用户界面,在所支持的操作系统间具有可移植性。

#### (23)CADAM IUE(CATIA-CADAM IUE)

CADAM IUE 是一个开放式体系结构工具,它可以充分发掘 CADAM 设计数据库的强大功能,以用于设计、分析、制造和管理过程的许多应用程序中。这些应用程序可以是可运行的第三方开发的应用程序,也可以是用该产品提供的一套完整的 FORTRAN 和 C 语言开发工具开发的应用程序。CADAM IUE 可在 CATIA-CADAM 制图环境下执行应用程序。同时具有完整的交互式用户对话功能。

#### (24)CPW/自动化测试工具(CPW/Automated Test Tools)

CATIA CPW/自动化测试工具产品是一套完整的计算机辅助软件工程(CASE)工具,用来记录、重放和比较一个交互式 CATIA 会话的结果。软件开发人员可以同时重放多个会话以进行功能、性能测定和回归测试。用户可以记录在一个平台和图形适配卡上的会话,进而在另一个平台和图形适配卡上进行重放。这些工具是 Dassault Systemes 用来开发 CATIA 应用系统环境的一部分。这些应用程序在所支持的操作系统间是可移植的。同时具有一致、高质量、可靠并持久的代码和数据。

#### (25)CPW/软件环境(CPW/Software Environment)

CATIA CPW/软件环境产品是一套完整的计算机辅助软件工程(CASE)工具,用以开发基于 CATIA 的应用程序。这些工具方便了在编译和交互式会话过程中检查错误,并测算在测试过程中有应用程序的百分之多少经过测试,确定在性能优化过程中 CPU 的使用情况。这些工具是 Dassault Systemes 用来开发 CATIA 应用系统环境的一部分。这些应用程序在所支持的操作系统间是可移植的,同时具有一致、高质量、可靠并持久的代码和数据。

### 4.2.8 AEC 工厂设计

#### (1)AEC 基础框架

CATIA-CADAM AEC 基础框架可以快速开发并优化 AEC 应用程序,采用面向对象技术在模型中代表

部件(如管道、梁、泵)、相关信息和智能。该产品具有的主要功能有 AEC 基元和线架通用 3D 造型工具,基于约束的参数化造型器,可客户化的报告生成器,这些功能可使用户或第三程序开发部门可以存取 AEC 对象。

#### (2) AEC 传送装置设计

提供功能强大的三维设计环境,用户可中库从调用多种部件,并带有属性特征。本产品多有方便的放样及编辑工具,并可实现从二维布置图到三维放样的自动转换,提高了设计交率。AEC 传送装置设计提供了多种干涉检查工具,如静态、动态检查等。同时,用户可自定义计算公式来预测费用、如送量、传送力等。本产品还提供与传送仿真软件 Deneb QUEST 的数据接口。

#### (3) AEC STEP AP227 接口

该接口采用国际标准“STEP”实现数据交换。同时,该接口还可转换厂房设计中空间信息,这些空间信息的类型由 AP227 协议定义。用户可从 CATIA 中双向地生成或读取遵守 STEP 格式的文本文件。该产品符合 ISO 标准,可实现不同计算机辅助系统间的数据交换,支持厂房设计包括操作及维护在内的整个生命周期。

#### (4) AEC 虚拟产品管理

CATIA-CADAM AEC 虚拟产品管理在集成部件几何数据和工艺数据的部件数据库中管理项目信息。它在全生命周期内支持 AEC 项目,从概念设计到详细设计、制造和施工到走动、运行、维护和拆除。用户可在并行工程环境中对同一部件的不同表达(草图、3D 设计、图样)进行操作,该产品采用了一个标准的 SQL 数据库仓库,并提供先进的修改控制功能。

#### (5) AEC 结构 FEA 集成

CATIA-CADAM 结构 FEA 集成可使用户将第三方和常用结构分析求解器集成到 CATIA-CADAM 工厂设计环境中来。在一个一致性设计和分析的环境中提供这些工具使得用户不必进行不同系统的培训,并在设计和分析任务间提供完全的集成。结构 FEA 集成对表、规则方法和其它方面的完全客户化提供了开放式体系结构。

#### (6) AEC 结构框架分析

CATIA-CADAM AEC 结构框架分析可快速地对用结构框造型产品建立的框架结构进行静态分析。用户可以直接对其 3D 模型进行操作来定义边界条件和载荷,系统随后自动地生成一个理想的分析模型。该集成化的求解器将生成应力,包括弯矩、剪切力以及反作用力等。分析结果以几何形式叠加在 3D 模型上,以增强用户的理解。

#### (7) AEC 结构框架造型

CATIA-CADAM AEC 结构框架造型为用户提供丰富的工具来构造 3D 框架结构。开发、扩展、革新或维护一个加工厂和发电厂、大型商业和工业建筑、轮船或其他结构的设计工程师可以提高设计质量和生产率。系统可以根据用户定义的规则和参数,可放置并连接符合国际标准的结构成员。项目管理员可以用建立功能来建立设计标准,以确保一致性并符合公司和政府的规定。

#### (8) AEC 制图

CATIA-CADAM AEC 制图对建立和管理 2D 相关图样提供全功能制图工具。采用透明模式,系统可以自动地由 3D 设计模型建立图样。这些 3D 设计模型是一些正交视图。需要生成基本施工图样和辅助图样或全标注和全尺寸标注平面图、剖面图或结构设计的工程师,可从制图产品的正文、尺寸标注、修饰和浏览功能中收益非浅。

#### (9) AEC 图样生成

CATIA-CADAM AEC 图样生成产品可由 3D 工厂模型自动生成符合公司制图标准的 2D 管道、设备和结构工程图样。用户可以快速建立并规划平面图、断面图、剖面图和轴侧图,将视图剪裁到棱柱区。设计并建立一个加工厂或发电厂所需的图样数量有几千个,通过自动化建立准确、可变比例、标注尺寸并加有标注的 2D 图样,图样生成产品可以减少设计和建立一个加工厂或电厂所需的工作量。由于 2D 图样是由 3D 模型直接生成的,用户可以确信文档与 3D 模型是完全一致的,并且通过相关性,在模型与图样任一编辑后,同样会保持一致。

#### (10) AEC 设备设计

CATIA-CADAM AEC 设备设计产品建立并管理用在发电厂、加工厂和造船厂的智能化设备和接头表达。它同 CATIA-CADAM 工厂设计应用产品完全集成。在一次操作中,用户可以很容易地对来自用户定义库中的全定义设备进行放置和定位。通过技术要求清单或借助于综合性的工具,它也可以参数化,系统允许用户修改对象定义,包括几何图形、数据、属性等。

#### (11) AEC 输送管设计

CATIA-CADAM 输送管设计产品允许用户自动进行整个输送管设计过程。便捷地对加工厂、发电厂和造船厂所用的参数化输送管零件进行放置和定位。输送管设计是一个集成化产品,擅长进行输送管设计工作;建立、修改、生效、图样抽取以及强大的编目生成。象所有 CATIA/CADAM AEC 工厂设计产品一样,输送管设计在各个专业学科间完全集成化,例如,管道、结构以及设备,可沿多专业的生成对象铺设输送管,智

能化输送管放置是全参数化的,并遵守缺省和(或)用户技术规范和标准。

#### (12) AEC 管道设计

CATIA-CADAM AEC 管道设计提供了一系列广泛的工具,用于对电厂和加工厂等管道系统的设计、修改、文档编制和管理。管道设计可使整个管道设计过程自动化。由一个大型 ANSI 管道编目,启动表和零件编目开始,管理员可以建立项目技术要求和标准,以确保设计质量。基于一般行业规范的智能化管道铺设、零件放置和编辑等,使工程师可建立一致性的、优化的设计。

#### (13) AEC 管道和装置图

CATIA-CADAM AEC 管道和装置图产品提供了用于电厂、加工厂和制造厂的管道和装置及相关数量、部件抽取的设计、修改、文档和管理的工具。该产品同时有智能可客户化库。这些先进的功能确保了高生产率以及与 3D 的集成,包括零件的智能化放置、线和部件的快速修改。该产品也提供了一套综合性选项,来对系统的行为进行几何可视化显示。

用于管道系统的设计、修改、文档编制及管理,使整个管理设计过程自动化。规则驱动的自动附件放置;P&ID 驱动的智能管路放样及附件放置;快速建立管路规格及尺寸表自动的附件一致性检查;强大的管系标准库建立和管理。可同时多个文本方式的设置表格输入到系统中;输入(输出)标准件库定义从(到)一文本文件中;可同时调用多个标准件库;在进行管系调整(如 push,relimit 等功能)时,可设置定位基准点;可将管路截为多段,每段的端点均可根据设计的要求而修改。

#### (14) AEC 管道轴侧图

CATIA-CADAM AEC 管道轴侧图生成是一个高度自动化系统,用于加工厂和电厂的管道系统相关轴侧图的快速生成。一个典型的工厂设计项目可以生成数千张轴侧图,用该产品生成轴侧图比手工建立大大地节省了时间,轴侧图的手工建立要消耗整个管道设计工时的 30%。系统生成可客户化的材料清单功能进一步提高了用户的工作效率。

#### (15) AEC 船体设计

CATIA-CADAM AEC 船体设计产品为各种船只船体结构建立提供了一个专用应用产品,它可生成一个船体的 3D 相关模型,该产品是对类似于 CATIA-CADAM AEC 图样生成产品这样的其它 CATIA 产品的一个完善,为准确、灵活并高效地船舶设计与生产提供了一套无缝集成化工具,造型是基于 CATIA AEC 造型器,并与这些造型 100%的一致,可提供无与伦比的跨专业集成。

## 5 AUTOCAD2004

### 5.1 AUTOCAD2004 概要

AutoCAD2004 是理想的二维制图及详图和三维设计工具,可加快数据创建,让您更加轻松地共享文件。AutoCAD<sup>®</sup>2004 软件引入了全新的功能,如生产力工具和演示图形,可使您更快地创建数据;CAD 标准工具和 DWF<sup>™</sup>(Design Web Format<sup>™</sup>)格式的电子文档,能让您更加轻松地共享数据;以及新的软件许可工具,可使您更加高效地进行数据管理。而且,它是迄今最快的 AutoCAD 版本。无论您使用 AutoCAD 进行三维制图、详图和设计文档制作,还是进行说明性三维设计,都可以对该软件进行个性化处理或编程,以满足您的特定设计需要,或添加能与 AutoCAD 配合的特定行业应用程序。AutoCAD 软件是计算机辅助设计(CAD)的全球标准。AutoCAD 的用户超过了其他任何 CAD 软件的用户。AutoCAD2004 是理想的二维制图及详图和三维设计工具,可加快数据创建,让您更加轻松地共享文件。

### 5.2 AUTOCAD2004 特点

#### (1) 更快地创建设计数据

在执行日常设计任务时,速度就是一切,如打开和发送文件、编辑标注、制作演示图纸和访问所需的工具等。这些新的 AutoCAD2004 增强功能使您能够更快、更加有效地创建设计数据。

1) 通过新的 DWG 格式获得高速体验:AutoCAD DWG 文件进行了优化,比运用旧版软件创建的文件小 52%。这意味着在通过电子邮件发送、上传和下载文件时可大大缩短文件打开和传送时间。

2) 运用新工具提高生产力:新的 AutoCAD 工具面板对于清理屏幕空间和提高生产力发挥了重要作用,这些工具面板的透明度可以调整,能够增大屏幕工作区域,并且可以充分进行定制,因此您可以将日常使用的内容保存在一个方便的位置。例如,您只需从工具面板将图块拖入您的图纸即可,而不必使用命令插入它。Express Tools(包括图层管理、尺寸标注和对象修改)减少了完成工作所需的步骤。运用更新的“重做”功能,您可以跟踪修改历史,恢复多次“撤消”操作。而且,第一个 AUGI 希望列表项目(多行文本)现已包括定位点和缩排功能,并且删除了“文本编辑”对话框,可提供更加用户友好的体验。

3) 新的演示图形:现在,您可以运用 AutoCAD 应用程序所包含的高质量图形制作演示图纸,而无需额外的软件。在两种颜色或同一颜色的明暗色彩之间指

定梯度填充。运用描影 Viewport 出图功能,打印演示质量的描影、三维等角视图。而且,通过 1600 多万种可供选择的 24 位真彩色,包括 PANTONE<sup>®</sup>、RAL CLASSIC 和 RAL DESIGN 颜色系统库,您可以向 AutoCAD 对象应用自己想要的颜色。

### (2) 更加轻松地共享设计数据

许多人可以同时进行一个设计项目,包括承包商、分包商、业主和工程师等,这里不再一一列举了,而且每个人都有不同的视角。但是,无论一个团队是多么千差万别,所有人都朝着一个共同的目标努力:成功的项目。而且为实现这一目标,他们需要交换信息。新的 AutoCAD 2004 功能使您能够比以前更加轻松地共享数字设计数据。

1) 安全地共享数据:通过新的密码保护、数字特征文件和增强的 DWF 文件格式,可使您安全地共享数字设计数据。您可以使用密码保护来确定哪些人能够打开您的文件。数字特征文件墨水签字的能效相同:验证您图样的来源、真实性和未修改状态。而且,使用 DWF,您可以通过英特网与需要查看和出图而不是编辑您的 AutoCAD 图样的团队成员交换图档文件。DWF 提供公支持查看和出图的锁定轻型格式,可生成与 DWG 文件相同的按比例视觉保真度。而且,您可以将多幅图纸发布为单一 DWF 文件,以简化传输。

2) 共享内容:如果图样已经绘制出来,请不要浪费时间重新绘制,使用它即可!您可以从更新的 AutoCAD<sup>®</sup> DesignCenter<sup>™</sup> 直接将现有的设计内容(如图块、标准、布局甚至整个 DWG 文件)拖入您的图纸。也可以使用新的选项卡访问 DesignCenter Online(访问大量预绘制 i-drop<sup>®</sup> 内容的接入点),您可以简单地从 autodesk.com 或参与厂商的网站将内容拖入您的制图会话,而没有下载、保存和插入命令的烦扰。厂商也可以在一个方便的 i-drop 文件包中附加关联的设计信息,如电子表格和订单。

3) 共享标准:运用 AutoCAD 2004,您将不必检查标准。该软件已具备“标准意识”:在您进行制图时,Standards Manager 会自动在后台运行。当违反或违背某个标准时,它会显示一条消息,并立即通知您违反了标准,提出纠正建议。而且,您可以选择要检查的标准类型(如尺寸和文本样式、图层或线型),并以读者友好的格式通过电子邮件发送或打印标准核查报告。

### (3) 更加有效地管理软件

AutoCAD 2004 软件具有众多新的工具,可使您有效地管理和获得最大的技术投资回报。例如,运用 Autodesk Product Manager,您可以从单一位置跟踪多个软件许可的版本、序列号和 PC 编号,而不再需要

访问每个办公室的每台 PC。如果您的设计人员需要在途中工作,可以通过 Autodesk Network License Manager(NLM)方便地从您的网络借用软件许可。他们可以提前归还许可,也可以等待指定的借用时限到期。无论哪一种方式,NLM 都会自动在您的服务器上续借许可。

## 5.3 功能与特性

### (1) 打开文件和保存文件

无论图形存储在本地还是网络驱动器中,都可以节省访问图形的时间。如果 AutoCAD<sup>®</sup> 2004 DWG 数据位于网络驱动器上,则打开图形和保存图形的速度分别比 AutoCAD<sup>®</sup> 2002 快 30% 和 66%。

测量只是初步的结果,它基于在受控网络上进行的自动测试。测量结果是近似值,可能会出现错误并进行更改。产品信息和规格如有更改,恕不另行通知。Autodesk, Inc. 以原样提供此信息,不作任何形式的明确或隐含的保证。

### (2) 制表位和缩进

多行文字(mtext)编辑中现在包含制表位和缩进。因此可以轻松地创建段落,并可以轻松地相对于文字元素边框进行文字缩进。

### (3) 外部参照通知

多个用户可以使用同一个图形集,以确保他们使用的是最新图形。如果外部参照已被更改或保存到当前图形,用户将在状态栏中收到实时和图标警告。

### (4) 多页 DWF

使用新的 DWF<sup>™</sup> 6(Web 图形格式)文件格式,可以将图形集中的所有图形发布到单个 DWF 文件。尽量减少传递的文件数量,可以避免在查看和打印文件时,文件顺序出现混乱。使用 Autodesk 的免费 DWF 查看器 Autodesk<sup>®</sup> Express Viewer,可以轻松地按顺序查看和打印多页文件。

### (5) Autodesk Express Viewer

Autodesk Express Viewer 是小型、快速的免费查看器,其中包含一些便于使用的工具,可供任何需要按比例查看和打印 DWF 文件的用户使用。设计信息的创建者和使用者可以轻松地访问设计图形,以便通过电子方式查看或进行打印,无需使用 AutoCAD 软件。了解更多信息。

### (6) 真彩色

用户可以在 1600 万种颜色中进行选择,并将所需的颜色应用到 AutoCAD 对象中。使用真彩色(24 位色)指定颜色设置;可以使用“色调”、“饱和度”和“亮度”(HSL)颜色模式指定设置,也可以使用“红”、“绿”、“蓝”(RGB)颜色模式指定设置。

颜色系统库包括 PANTONE<sup>®</sup>、RAL DESIGN 和 RAL CLASSIC。

\* RAL DESIGN<sup>®</sup> RAL, Sankt Augustin, 2002; RAL CLASSIC<sup>®</sup> RAL, Sankt Augustin, 2002。RAL 颜色的表示形式得到了 RAL Deutsches Institut für Gütesicherung und Kennzeichnung e.V. (RAL-German Institute for Quality Assurance and Certification, reg. Assoc.), D-53757 Sankt Augustin 的批准。

请注意,显示器上的颜色显示只与注册为着色样本的实际颜色着色相似。当前所用软件或硬件的质量均与此无关。

#### (7) 渐变填充

使用新增的“渐变填充”功能,可以直接在 AutoCAD 应用程序中创建演示图形,而无需使用其他软件。指定一种渐变填充:在两种颜色之间进行平滑过渡,或者使用一种颜色由深到浅地平滑过渡。

#### (8) 联机设计中心

通过 AutoCAD<sup>®</sup> DesignCenter<sup>™</sup>功能中的新选项卡,可以从单个位置访问本地驱动器、网络驱动器或 autodesk.com 上的内容。该选项卡中包括直接指向数以万计的符号库和 Autodesk.com 上的制造商内容的链接(均为 i-drop<sup>®</sup>格式)。

#### (9) 口令保护

通过向图形文件应用口令,确保未经授权的用户无法打开或查看图形。

#### (10) 许可证借用

Autodesk<sup>®</sup>网络许可管理器(NLM)允许用户在有限的时间内借用网络许可证,并在重新连接至网络时将其返还至服务器。使用客户端安装的 AutoCAD 软件版本,NLM 用户可以在不访问网络许可管理器的情况下临时工作。

## 6 MDT 软件

### 6.1 MDT 软件概述

MDT 软件(Autodesk Mechanical Desktop)集零件造型、曲面造型、装配造型和自动绘图等于一体,是一种面向现代化机械工程设计的三维设计工具集成软件包。

对于使 AutoCAD 的设计者来讲,熟悉 MDT 的界面和工作环境是很容易的事情。它可以将设计者由手工绘制平面工程图的过程变为通过三维参数化实体造型进行产品设计的过程。MDT 允许设计者根据需要随时对各种造型进行更改,并能自动将与其相关联的内容全部刷新。

它主要包括 AutoCAD Designer, AutoSurf, AIT, Autodesk WorkCenter 和 AutoVision 等。其中,主要的 CAD 功能核心是 AutoCAD Designer 模块和 AutoSurf 模块。它可以帮助设计者把要设计的机械产品首先实现概念化,然后再对其进行造型设计,最后生成产品文档。MDT 的每一种工具都提供了一种特定的设计功能,如实体造型、曲面造型或装配造型等。MDT 集这些功能于一体,提供给设计者一个强有力的机械设计综合工具。

综合使用 MDT 的各种工具完全可以满足一个复杂设计问题的特定要求。设计者可以自如地运用各种工具表现自己的设计意图。这种特性使 MDT 与那些迫使设计者要遵循非常固定的设计变量表的软件包相区别,使 MDT 成为一种真正是为设计者服务的设计工具。

MDT 提供了强有力的参数化设计方法。参数化设计方法允许设计者以零件的尺寸、几何形状和相互位置等关系来进行产品造型设计。MDT 能识别这方面的关联信息,并对造型的任何修改在整个设计中可以方便、快捷地反映出来。

MDT 的另一个优点是,它运行在大众化的 AutoCAD 环境下,因此对熟悉 AutoCAD 的设计者来说,这显然会减少学习的困难。

### 6.2 MDT 的功能介绍

#### 6.2.1 AutoCAD Designer 模块的零件造型功能

常见的机械零件的结构大多是由很多棱柱形物体组合而成的。对这一类零件的设计工作完全可以通过已经集成好的 AutoCAD Designer 的零件造型和部件造型功能很方便地实现。AutoCAD Designer 包括 Part 和 Assemblies 两个子功能模块,是一个具有零件和装配双重造型功能的三维参数化实体造型工具。设计者可以用 AutoCAD Designer 完成对各种零件的造型,然后再把它们组合成复杂的部件或装配模型。

使用 AutoCAD Designer 的 Part 子模块,设计者通过绘制零件各部分形状的草图,定义它们的尺寸、位置和内在的相互关系等步骤来进行零件设计。对绘制的零件的这些形状又可称其为“特征”。利用种种“特征”就可以像堆积木那样来构造零件的各种复杂形状。

因为 AutoCAD Designer 对零件的这些“特征”设计都是采用参数化设计方法,所以,如果改变其中一个“特征”的形状或位置结构,AutoCAD Designer 能够自动地改变零件中所有与其相关的各种“特征”以便反映这个变化及其带来的影响。



作为 MDT 中一个相对独立的工具, AutoCAD Designer 可以设计各种棱柱类零件, 这类零件在各种设计中通常要占很大的比例。

MDT 为了满足各种类型物体造型的需要, 还允许组合使用 AutoCAD Designer 和 AutoSurf 的造型功能以实现零件进行各种形状的造型设计。例如, 可以用由 AutoSurf 构造的曲面来切割由 AutoCAD Designer 生成的实体以创建满足设计要求的复合形状的结构等。

### 6.2.2 AutoSurf 模块的曲面造型功能

在设计冲压、模锻、铸造类产品或铸锻模具时, 软件的曲面造型功能是非常重要的。AutoSurf 是一个能够为组合型物体创建复杂模型的三维曲面造型工具。用它可以设计各种用自由曲面来表达的轮廓或雕塑曲面, 如在汽车上经常可以看到的各种内部的或外部的曲面形状。

作为 MDT 中一个相对独立的工具, AutoSurf 既能满足常规的曲面造型需要, 也可以根据原始数据或已数字化的物理模型(如泥塑模型)的截面进行线框式曲面造型。AutoSurf 模块还能通过为数控设备和刀具路径提供数据的方式实现 CAM 过程。

作为 MDT 的一个组成部分, AutoSurf 还可以通过修改别的 MDT 工具(如 AutoCAD Designer)创建的实体来进行曲面造型。AutoSurf 允许设计者借用原有的一些模型元素来组建新的复合形状的机械零件曲面模型。

AutoSurf 构造的曲面模型基于非均匀有理 B 样条(Nurbs)曲线。AutoSurf 通过控制 Nurbs 曲线的线框式模型来构成不同类型的曲面。

### 6.2.3 AutoCAD Designer 模块的装配造型功能

在创建好零件模型后, 利用 AutoCAD Designer 的 Assemblies 子模块可以把各个零件模型组装起来, 来生成各种大型的部件模型或大型的装配模型。AutoCAD Designer 使用与零件造型中相类似的装配造型变量表来进行装配造型的识别, 只是在进行装配造型时组织的是“构件”之间的相互关系, 而不是零件造型时采用的所谓“特征”之间的关系。在装配模型中, “构件”既可以是单个的零件, 也可以是装配成一个部件的一组零件的集合。AutoCAD Designer 可以根据各个“构件”的配合与联系关系, 把诸多的“构件”按照一种逻辑关系组装在一起形成一个装配模型。

像零件造型一样, AutoCAD Designer 能对这些“构件”的配合与联系关系进行参数化识别, 而且每当一个“构件”发生任何改变时, 它都能够迅速识别并刷

新整个装配模型。

### 6.2.4 设计文档

通常, 工程图样与各种文档资料就是真正的设计文件, 因为人们根据它们就可以指导机械设备进行产品的加工制造。由于 MDT 的各种工具都包含在大众化的 AutoCAD 环境之中, 所以 MDT 可以方便地获得 AutoCAD 全部的绘图功能。但 MDT 生成设计文档的功能却是自带的。

AutoCAD Designer 和 AutoSurf 为生成工程图样提供了强有力的支持, 能为设计者完成各种工程图纸的创建工作。这些工具根据零件造型或装配造型, 能够自动创建传统的二维正交图、轴测图、辅助图、截面视图及零部件的各种局部视图等等。

AutoCAD Designer 和 AutoSurf 可以根据实体模型的实际尺寸自动生成带尺寸标注的工程图样, 也可以在工程图样上添加新的注释或增加其他必要的尺寸。因为各个视图都是根据实体模型自动生成的, 所以, 当设计者对某个模型进行修改时, MDT 能够自动刷新所有的有关视图。MDT 提供的这些绘图功能可以为设计者节省大量的绘图时间, 而将精力集中在统一、完善必要的文档资料上, 以更好地支持生产制造过程。

### 6.2.5 图形转换

使用 MDT 进行工作的设计人员可能还想使用别的 CAD 工具来辅助 MDT 的工作, 为此, 在 MDT 内部带有一个 IGES (Initial graphic exchange specification) 文件转换器, 即 AutoCAD IGES Translator, 简称 AIT。IGES 作为可以在很多 CAD 系统之间进行数据交换的标准, 得到大多数 CAD 销售商的支持。在 MDT 中使用的 AIT 遵循最新版本的 IGES 和相关标准。

AIT 除了可以创建和维持一个柔性的 CAD 工具环境外, 它还支持和保护过去用别的 CAD 系统开发的产品图形的数据。AIT 支持下列各种类型的设计对象:

- 1) 各种二维和三维线框式几何图形;
- 2) 规则曲面、参数化曲面和非均匀有理 B 样条(Nurbs)曲面;
- 3) 用 AutoCAD Designer 和 AutoCAD 系统创建的各种实体, 以及用 IGES 文件表示边界的实体。

由于参数化实体造型方法对 CAD 系统来讲是相对较新的功能, 而设计者过去设计的大多数图形文件都是以非参数化的二维和三维实体形式保存下来的。为了解决这个问题, 在 MDT 中可以通过 IGES 转换器



(即 AIT), 把过去所做的设计工作转换到 MDT 中加以保存, 并可以进一步对其进行参数化处理。例如, 把 AIT 和 AutoSurf 结合使用, 可以对在 AutoSurf 中创建的曲面进行再处理, 然后将其与参数化的实体模型组合起来。这时, 在 AutoCAD Designer 中就可以获得一个取代旧设计的新的参数化实体, 而且它将保留着在 AutoSurf 中最初创建或修改过的复杂的曲面形状。

### 6.2.6 Autodesk 的相关设计工具

把 MDT 与 Autodesk 公司和它诸多的 MAI (mechanical application initiative) 合作伙伴的工具结合起来使用可以极大地扩展它的设计功能。MDT 在这些工具的辅助下可以创建一个能够满足设计人员特殊需要的 CAD 环境。

除了 MDT 以外, Autodesk 公司还提供了其他一些可以帮助设计人员进行工程管理。提高工作效率和增强设计质量的工具。其中有:

#### (1) 文档管理工具 Autodesk WorkCenter

在产品的设计过程中, 设计人员需要随时随地地创建并保存各种文档和图形资料。一个好的文档管理系统可以使设计者非常方便、安全地完成这些工作。与 MDT 融为一体的 Autodesk WorkCenter (以下简称 WorkCenter) 是一个可以对技术文档进行分类管理的文档管理系统。它可以完成对 CAD 图形文件、各种规范、数据列表、数据库等的分类管理, 使识别、定位、检索、修改、共享和安全储存等操作更加方便, 同时它还具备对工作流程的自动处理功能。

WorkCenter 与 MDT 的装配造型功能紧密集成在一起。MDT 的 AutoCAD Designer 模块除了可以完成零件造型与装配造型等设计任务外, 还可作为检测工具来监视和管理各种设计模型。如利用它的装配造型功能, 可以把设计好的装配模型很方便地分解为各种部件模型和零件模型。这些构件模型不仅可以在本设计中重复使用, 而且还可以通过外部引用绘图文件的方式在不同的设计模型中重复使用。所有这些都是通过 WorkCenter 在 MDT 内部自动进行管理的; 再者, 当设计者需要对设计文档内容做出修改时, WorkCenter 可以自动实现相关的变化来帮助设计者进行这方面的管理。

#### (2) 机械数据库

Autodesk 的机械数据库可以为机械设计和制造提供期刊式的智能数据服务。它以 CD-ROM 光盘的方式定期为设计者提供不断更新的、并能进行迅速检索的各种机械数据库。

Autodesk 的机械数据库将各种必要的机械信息

集中放在一起, 完全可以淘汰各种散乱的目录和手册。这种机械数据库每半年更新一次, 以保持最新的机械信息。在 Autodesk 的机械数据库中, 提供了帮助设计人员迅速、方便地查找所需数据的智能接口。

Autodesk 公司提供的可与 MDT 配套使用的机械数据库有以下两种:

1) 零件数据库 PartSpec: 在设计过程中, 设计者大都经历过这样一些乏味的中断, 如翻阅各种技术手册和产品目录, 重新设计一个曾经使用过的通用零件等。为了很好地解决这个问题, Autodesk 公司提供的零件数据库 PartSpec 收集了二十多万个基本零件和 2 亿多个标准联结件。对它们都可以通过简单的“拾取和安放”操作迅速地直接插入到设计文档中, 而且它们在插入后仍会保持各自的属性信息, 从而可使设计者把更多的精力放在真正的设计上。

PartSpec 收集的零件图形文件主要来自以下零件生产厂家: Adept Technology, Boston Gear, Jergens, Parker-Hannifin 和 Timken 等。

2) 材料数据库 MaterialSpec: 通常, 设计中选择材料的工作也会使设计、分析工作放慢速度。MaterialSpec 是一个强有力的、可交叉参考的工程材料数据库。它储存有大量的各种工程材料数据, 包括塑料、陶瓷、金属和复合材料等。它全面提供了世界上 300 多家生产商的 25000 种材料信息, 可使设计人员迅速地查寻到合适的工程材料并运用到设计中去。

使用 MaterialSpec 非常方便、简捷。每条记录都包含有该材料的有关材料特性、所属类目、适用范围等方面的完整信息。当设计者确定对材料的基本需要后, 就可以在材料数据库中方便、简捷地查找到满足要求的材料。

#### (3) 设计渲染工具 AutoVision

AutoVision 是一个可渲染 MDT 中创建的三维造型的演示工具。使用 AutoVision 能逼真地展示设计人员的设计概念, 有助于直观预视设计者所设计的产品。

过去, 要实现对设计内容的人工渲染只能运用水彩、蜡笔、油墨等材料及喷涂设备对设计对象进行着色处理, 以表现最终的外观效果。而 AutoVision 则综合运用几何形状、灯光效果和设计对象完整的表面信息, 可以把产品渲染得非常逼真。AutoVision 的渲染设置方式有缺省设置和高级设置两种, 以满足设计者不同的需要。

#### (4) 动画制作工具 3D Studio

3D Studio (简称 3DS) 是一个具有扩展的渲染功能和动画制作能力的三维造型和动画制作工具。运用 3DS 的渲染功能, 能够按照多种工业标准格式提供高分辨率、真实感强的演示、印刷和其他用途的高质量

图像。

运用 3DS 的动画功能,可以对所设计的机械产品进行运动仿真。通过仿真可以预测所设计产品的工作过程及其机器的外形。使用 3DS 可以对演示对象进行移动、比例缩放、旋转、改变物体的形状和材料等精确控制;还可以模仿布光、摄影机、聚光灯和照相相对焦等手段来产生电影效果。

### 6.2.7 Autodesk MAI 合作伙伴提供的机械工具

Autodesk 公司同开发机械分析工具和加工制造工具的诸多第三方植智商建立起一种合作伙伴关系,这种合作伙伴关系被称为 Mechanical Application Initiative (简称 MAI)。MAI 合作把 MDT 同一些用于 CAD 特殊分析领域的开发工具,如造型分析、有限元分析、公差分析、运动学分析等集成起来。

另外,MAI 合作还把 MDT 与许多可以实现刀具路径控制和数控设备控制的工具结合起来。这类设备包括冲床、车床、铣床和磨床等。这些工具允许以自由格式曲面或实体模型等方式控制刀具路径和加工过程。

MAI 合作也支持那些以高效率工具对 MDT 进行专业开发或功能扩充的销售商。允许它们在特殊符号库、参数化结构特征、装配造型扩展、材料明细表、SQL (structure query language) 连接及其他方面对 MDT 的功能给以增强。

MAI 合作使 MDT 和其他 MAI 伙伴的工具实现高度集成。MAI 伙伴的各种工具之间具有通用的用户接口,它使 MAI 伙伴的各种工具非常易于被用户连接与熟悉。这个集成还包括对各种数据库的集成。在 MDT 和 MAI 伙伴的各种工具之间可以实现参数化设计数据的双向通信联系。这些数据都能被 MDT 或 MAI 的其他工具直接识别和编辑。因为在这些工具之间完全不需要进行麻烦的数据转换,所以使用这些工具能迅速、精确地完成对设计数据的各种处理。

## 7 Solid Edge

### 7.1 Solid Edge 概要

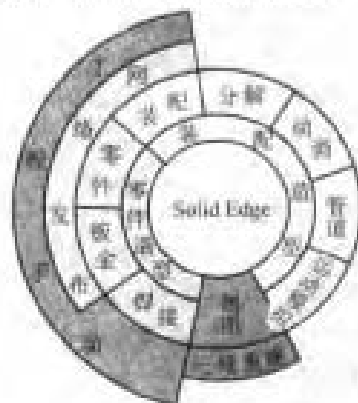
Solid Edge 是 EDS 公司的中档 CAD 系统,它以其卓越的性能、优异的造型功能、众口皆碑的易用性和专业化的设计环境赢得了业界广泛的赞誉。拥有 Solid Edge,您就能在第一时间设计出完美的产品,让您在竞争激烈的市场上占尽先机。

Solid Edge 是为机械设计量身定做的 CAD 系统,从零件设计、装配设计到工程制图,各种功能无所不

在。装配造型无以伦比,通用零件造型功能强大,专业化的钣金、管道、焊接设计独具特色,而制图模块则简洁明了。

不仅如此,Solid Edge 还提供了许多高效、独特的工具。网络发布让您点击几下鼠标,就能使产品在网上发布;工程手切设计、计算和造型于一身,使标准零件设计易如反掌;各种监测工具让零件缺陷无处藏身;运动模拟具有仿真功能,零件运动与实际机器毫厘不差。总之,您所需要的,正是我们所提供的。帮助您提高效率,完善设计并降低成本是我们多年来孜孜以求的目标。

Solid Edge 采用了 STREAM 技术,将逻辑推理和决策分析融入到机械设计的各个过程中,动态捕捉您的设计意图(图附-1-5)。动态工具条独具匠心,不管您身在何处,它都能为您提供反馈,引导您达到目的。各种命令的设计简洁清晰,并与实际加工方法紧密相关,使得操作过程自然流畅。您无需牢记命令的细节,就能在动态工具条的引导下轻松设计而不会迷失方向。同样是机械设计,STREAM 技术能减少鼠标和键盘操作达 15%~57%,提高效率 36%。



图附-1-5

Solid Edge 采用 Parasolid 作为软件核心,将中端 CAD 系统与世界上最具领先地位的实体造型引擎结合在一起,功能强大,是从事三维设计的优秀 CAD 软件,同时还提供了从二维视图到三维实体的转换工具。您无需摒弃多年来二维制图的成果,借助 Solid Edge 就能迅速跃升到三维设计,这种质的飞跃让您体验到三维设计的巨大优越性。

使用 Solid Edge,您不仅是在建立新产品的三维模型,而且还能获得完成精确设计的知识。工程助手能帮助您快速评估各种设计方案,从而优化机器性能和可靠性、质量特性计算、设计参数监视器、运动分析、干涉检查和其他多种内置工具,帮助您捕捉和实现您的设计理念。

Solid Edge 不但是一个独特的 CAD 系统,同时又

是实现系统集成的基础。基于 Solid Edge 的 CAM、CAE 软件丰富多样,且都能与 Solid Edge 无缝结合。通过系统集成,就能有效地剔除冗余数据,减少从设计到产品的开发周期。

一旦您选择 Solid Edge,您就加入了由专业销售商、产品开发商、技术支持工程师、培训专家以及其他相关用户组成的强大团队。您应为您的加入倍感自豪,因为您的背后是全球一流的团队,他们在不断地为 Solid Edge 及其用户提供强有力的支持。

## 7.2 Solid Edge 功能模块

### 7.2.1 Solid Edge 产品介绍

Solid Edge 采用多种技术,能够轻松完成大型机器的装配。采用自顶而下和自底而上两种装配技术,使装配设计可以在工作组中齐头并进,并确保整机器的正常装配。设计得当的装配方法和智能装配,提高了装配效率。显示配置能隐藏或冻结与当前工作无关的零件,既便于装配体又减轻了系统的负担。装配体族和多工位装配的引入有效地解决了装配体改型所面临的重新装配的繁琐问题。其他诸如装配分解图生成、动画文件制作、管道的建立都在独立的子环境中进行,设计巧妙而又简单实用。

#### (1) 装配关系和零件装配

Solid Edge 定义了贴合、对齐、插入、各种点线面的连接和夹角等装配关系,综合解决了零件装配的各种情况。装配零件的过程,实际就是定义零件与零件之间装配关系的过程。从零件库中拖动一个零件作为装配窗口,然后在动态工具条的引导下逐步指定装配关系就能完成零件装配。参照装配和自动装配命令能根据选择的目标自动判定适用的装配关系,简化了步骤,提高了装配效率。

#### (2) 装配管理器

装配管理器集中了用于装配的多种工具,包括装配树、零件库、装配族、监测器和搜索工具。装配树工具可以查看装配结构、编辑零件装配关系。零件库类似于 Windows 的资源管理器,可以浏览及装配零件。装配族能实现多种装配方案,通过零件替换就能在一个装配文件中反映不同零件的装配结果。监测器是实现精确装配的好帮手,能对指定的指标提供实时监测。搜索工具能迅速地查找并显示装配中的零件。与相应的浮动菜单相结合,使用装配管理器可以完成几乎所有的装配工作。

#### (3) 装配树

装配树工具用两个窗口显示装配信息。装配结果窗口列出了构成装配体的所有零部件,装配关系窗口

则显示当前零件的装配关系。结合浮动菜单,在装配结果窗口可以隐藏、显示、激活、冻结选定的零件,在装配关系窗口可以删除和抑制装配关系。

#### (4) 装配族

使用装配族工具,可以在同一个装配文件中保存一组仅仅有个别零件不同的装配体,为设计和评估各种装配方案提供了方便。

#### (5) 多工位装配

采用多工位装配,能够生成某个特定零件处于不同工作位置的装配体,为研究机器运动或制图提供了便利。

#### (6) 装配变量

装配环境的变量表,允许用户加入装配约束变量或参数变量,并可直接存取零件或子装配变量而无需临时打开相关的零部件。同时也可将装配体中不同零部件的变量关联起来,为装配定位提供了极大的方便。

#### (7) 阵列装配

对于用阵列方式建立的特征,在装配时可以以阵列方式快速地复制一个或多个零件或部件,并保持相关性。如果零件的特征阵列变化,装配阵列也能自动地随之改变。

#### (8) 智能装配

当零件的装配关系相同时,可以使用参照装配命令,让系统“记忆”零件的装配关系,并自动映射到新装入的零件中,从而提高装配效率。而自动装配命令则更是技高一筹,您无需指定,只要选择恰当的结合面,系统就能自动判定并施加相应的装配关系。

#### (9) 零件定位

在大型装配体中,若零件嵌套太深,定位或寻找它在装配体中的位置可通过零件定位命令轻松实现。只要在装配窗口中选择零件,然后右击鼠标,执行浮动菜单上的零件定位命令,特征树就会自动展开至选中的零件,它在装配体中的位置和相应的装配关系也就一目了然了。

#### (10) 装配管理

为便于管理大型装配体,可以使用零部件转移、零件归并、部件拆分和重新排序等装配整理工具。零部件转移能将零部件由一个装配体转移至另一个装配体而保持其装配关系。零件归并能够将选定的一组零件下浮一级,归并为一个新的子装配体。部件拆分能将子装配体拆散,并使其组成零件上浮一级。重新排序能在同一级装配结构中对具有固定装配关系或自由零件重新编排装配顺序。

#### (11) 内存管理

对于装配体中的零件, Solid Edge 提供了激活与冻结机制以降低系统负担,改善系统性能。用户可以根

据需要,认定冻结零件或激活零件。对被冻结的零件,系统仅保留用于显示的几何信息,而卸载其造型信息和数学定义,从而减轻了系统负担,轻松建立具有数以千计零件的大型机器。

#### (12) 零件显示方式

装配体中的零件可以显示,也可以被隐藏。显示或隐藏的零件可以是激活的,也可以是冻结的,使用隐藏功能,可以使与当前操作无关的零件从屏幕上隐去,既便于操作又改善了系统功能。

#### (13) 装配剖视图

装配剖视图是用一个假想的剖切面将整个装配体剖开后得到的一种图形。在装配剖视图向导的指引下排定剖切面位置,剖切方向和剖切范围后,装配体的三维剖视图就生成了。装配剖视图是说明装配结构的一种好方法。

### 7.2.2 Solid Edge 零件设计

Solid Edge 提供了参数化的、基于特征的造型工具,帮助设计师快速高效地设计零件。首先建立一个由旋转和拉伸造型生成的基础,然后再以增加材料或去除材料的方式建立其他特征。零件的建模过程与制造零件的实际加工过程一致。可以建立诸如开孔、除料、圆角、抽空等机加工特征,以及拔模斜度、扫描、扫掠、螺旋等复杂的几何特征或特征阵列。零件的尺寸、特征关系等都能够快速修改以反映设计方案的变动。

#### (1) 智能化设计技术

Solid Edge 提供了功能强大的智能化设计手段,帮助设计师实现其设计意图。

快速拾取功能能智能地判定拾取对象,若选择区有多个目标,则自动弹出可选对象编号,便于精确地选定目标。

智能导航功能能实时捕捉特殊几何点的智能工具,并用图标给以动态反馈。智能导航是实现精确定位不可或缺的高效工具。

动态工具条是根据当前命令动态改变的又一个智能工具,不管执行什么命令,动态工具条都能引导用户快速有效地完成设计。

#### (2) 功能强大的特征造型

Solid Edge 提供了种类繁多的特征造型命令,可以建立拉伸、旋转、扫描、筋板、螺旋、拔模斜度、薄壁、倒角、圆角、特征阵列和镜像等各种特征,为机械设计提供全面的造型工具。特征的建立方法方便易用,往往是简单几步就能大功告成,其效率和易用性众口皆碑。

#### (3) 参数化设计

参数化设计方法的采用,使得修改造型异常轻松。

只要改变造型参数,就能立即获得新的造型结果,为评估多种造型方案提供了方便。

#### (4) 特征管理

特征管理器是一种特征管理的综合工具,包含了特征管理、特征库、零件族、感应器和过程回放等工具。

特征管理工具具有特征选择、删除、更名、抑制、排序等特征编辑功能,结合动态工具条,几乎可以进行所有的特征管理工作。

特征库可以存放各种典型特征,以备造型时直接使用。使用特征库可以象搭积木一样快速地完成零件造型。

零件族是具有多个共同特征和个别不同特征的一组零件,可以通过特征抑制或增加特征来构造,零件族功能实现了由一个造型而派生出多种造型方案的目的。

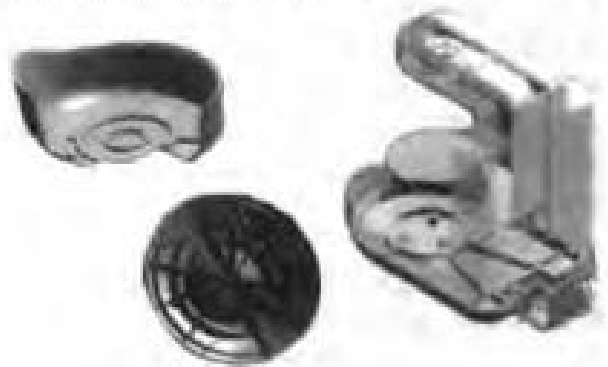
感应器可以为造型提供动态监测服务,过程回放能像播放动画一样展示零件的构造过程,是了解造型步骤或学习造型策略的理想工具。

#### (5) 塑料件造型

Solid Edge 提供一些强有力的造型工具用于实现塑料件造型,如凸缘、凹槽、零件分割、薄壁和网格筋板等。

#### (6) 质量特性计算

零件的质量、体积、表面积、重心、几何中心以及主轴方向和惯性矩等都可以由系统自动计算,并且相对的坐标系也可以改变(图附-1-6)。



图附-1-6

#### (7) 变量表

变量表类似数字表格,在变量表中,可以定义各种形式的变量,可以用公式来定义变量之间的数学关系,当然也可以通过改变变量的值,从而对相关的图形进行控制。

#### (8) 零件渲染

零件渲染不但可以对零件,而且可以对构成零件的边、面、特征等各层次要素赋予颜色或材质以便于观察。

### (9) 表格曲线和曲面缝合

使用 Excel 表格定义曲线上点的位置坐标,就能自动生成三维曲线(图附-1-7),曲线的类型及拟合方法也可以指定。拟合的曲面可以通过缝合生成单个连续光滑的曲面。



图附-1-7

### (10) 模型简化

模型简化可通过去除模型上的部分特征来达到简化模型,提高系统性能的目的。简化的零件与零件本身存储在一个文件中,特征管理器采用分节的方法进行管理。

### (11) 法向拉伸与除料

法向拉伸与除料允许沿曲面法向拉伸增加材料或去除材料。

### (12) 模型渲染

Solid Edge 的用户逼真渲染功能用户界面友好,操作简单,且效果令人叫好(图附-1-8)。用户可以通过指定渲染方法、灯光、背景和环境反射,定制物体所处的场景。而对于模型本身,还可以指定边的颜色、表面的颜色或纹理、凹凸纹理、外观特性和表面的反射特征,从而创造出一幅栩栩如生的画面。这些设置安排在一个统一的对话框中进行,点几下鼠标就可轻松完成。

### (13) 装配监测

在装配过程中,可以设置装配距离传感器,变量传感器或自定义传感器,用以监测零件之间的相对距离、装配变量或某特殊变量。这些传感器能实时提供反馈,帮助建立正确的装配体。

### (14) 干涉分析与测量工具

干涉分析是验证装配体工作时是否发生碰撞的理想工具。通过指定两个检查对象,并指定提供反馈的形式,系统就能自动分析判断零件是否干涉,并给出相应的反馈信息。若发生干涉,还可以使用系统提供的各种距离、角度等测量工具找出干涉原因。

### (15) 装配分解图

装配分解图用来反映装配体的零件组成、装配次序和零件的相对装配位置,是一种特殊的装配体。在 Solid Edge 中,有一个独立的装配子模块,用来生成这



图附-1-8

种装配分解图。分解图可以采用自动和手动两种方式产生。不过在一般情形下,只要单击自动分解命令,装配分解图就自动产生了,简单高效。分解后的装配体可以以一种显示配置保存在同一个装配体文件中,以生成装配分解的工程图。

### (16) 管道设计

管道设计也是在一个专用的装配子环境中进行的。在管道向导的指引下,通过指定管道起点、终点和管道的转折方位,就可自动生成连接管道。管道接头类型、转折半径及转折处的管接头都可以方便地指定。管道与其所连接的零件动态关联,若零件发生变动,管道也会自动更新。

### (17) 动画输出

动画制作模块是装配环境之下的一个子模块,可以快速生成全方位转动动画或由特定视图组成的关键帧动画。动画文件可以以图像或动画的格式输出,简单有效地达到了演示虚拟产品的目的。

### (18) 运动模拟

运动模拟是仿真机械运动的专用子模块,它也位于装配环境中。这个运动模拟模块是 MDI 公司的 SIMPLY MOTION。在运动向导的指引下,通过定义装配体中的固定零件和运动零件,各零件之间的连接副和主动运动零件的运动就能使模型运动起来,从而实现运动模拟及运动干涉分析。

## 7.2.3 Solid Edge 流水线式的制图模块

Solid Edge 是专为机械制图开发的,它提供了优异的图形表达、局部放大、技术标注和尺寸控制工具,

并能自动符合您选定的制图标准。无论是从零件模型、装配模型,还是一张空白图纸出发,Solid Edge 的制图和标注工具都能使您便捷地完成制图,它的性能超出了其他任何 CAD 系统。

#### (1) 视图向导

视图向导操作简单,但功能强大。针对不同的模型,比如零件、装配体、钣金件、焊接件等,视图向导会提供不同的选项。在视图向导的引导下,首先选择模型,然后指定主视图方向,最后指定基本视图的配置,就能得到一张准确的零件或装配视图,快捷方便。

#### (2) 向视图

向视图命令可以从任意一个基本视图出发,生成指定方向的向视图,并能自动地保持与参照的基本视图的对齐关系。

#### (3) 斜视图

斜视图命令可以从任意一个基本视图出发,按指定的投影方向生成斜视图。对于自动生成的斜视图,可以通过视图编辑,如旋转、隐藏部分线段等方法,生成具有特定表达功能的斜视图。

#### (4) 局部放大图

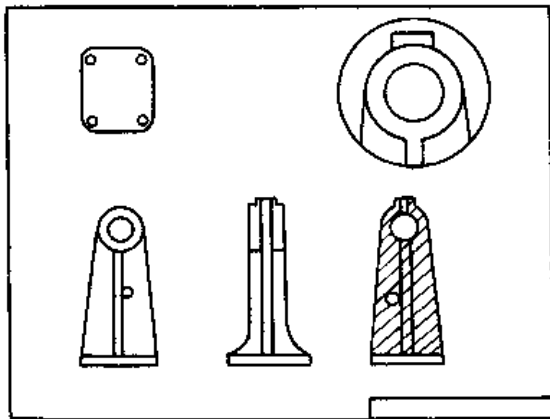
局部放大图命令可自动生成视图上某选定区域的放大图。局部放大图的比例及摆放位置均可自由设定。

#### (5) 断裂视图

对于结构简单但尺度较大的图形,可以采用断裂视图的表达方法,以便节约图纸空间。

#### (6) 剖视图和剖面图

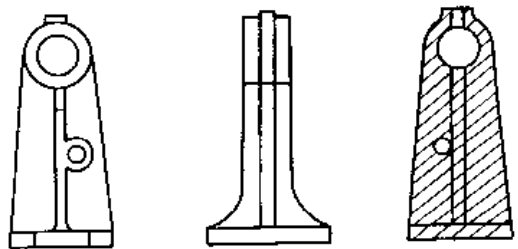
Solid Edge 的剖视图是由两个命令完成的。首先要使用剖切面命令在某视图上用画线或画弧命令画出剖切面的积聚投影,然后再使用剖视图命令由剖切面引出相应的剖视图。剖切面可以由一个或多个平面构成,生成的剖视图可以是阶梯剖视图、旋转剖视图、斜剖视图或复合剖视图(图附.1-9)。一个剖切面只能生成一个剖视图,删除剖切面将删除相应的剖视图。



图附.1-9

#### (7) 尺寸提取

Solid Edge 是参数化的特征造型系统,具有尺寸驱动功能,所以造型时一般要标注尺寸,以便对图形进行尺寸驱动。如果造型时使用了尺寸,就可以使用尺寸提取命令将造型时使用的尺寸自动标注在视图上(图附.1-10)。尺寸提取简化了尺寸标注过程,极大地提高了制图效率。



图附.1-10

#### (8) 智能标注

智能标注是一个通用的自动标注工具,它能自动判别指定图素的种类,并自动标出圆的直径、圆弧的半径、线段的长度等。这种将多种标注命令集成在一条命令中的方法,既简化了操作,又提高了效率。

#### (9) 模型与视图的双向关联

Solid Edge 自动生成的视图与模型之间是双向关联的。改变模型会导致视图发生改变,双击视图会自动打开模型到造型环境,但直接修改视图上标注的尺寸,不会使模型随之更新,相反,系统会给尺寸数字加下划线以反映视图与模型的差异。

视图与模型的关联关系可用视图转化命令把它们剖裂开来,这样的视图就转变为纯二维图形,并且这种转化是不可逆转的。

#### (10) 表面粗糙度、焊接符号及形位公差标注

Solid Edge 提供了标注表面粗糙度、焊接符号和形位公差的专用命令。通过设计完整的标注对话框,可以十分方便地完成表面粗糙度、焊接符号和形位公差的标注工作。

#### (11) 零件明细表

对于装配体,Solid Edge 可自动对其所组成的零件进行编号,并可自动生成零件明细表。零件编号可以用指定方式对齐。

#### (12) 孔表

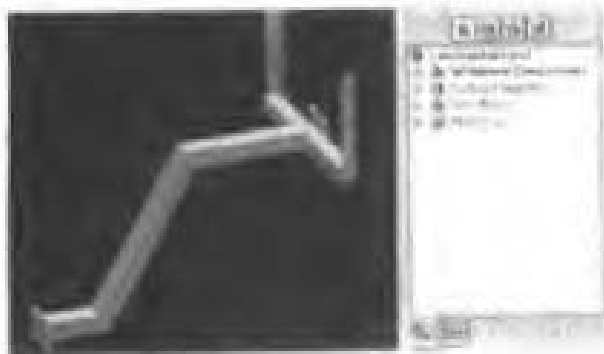
Solid Edge 能自动生成视图中的孔表,并可自动提取孔的深度。

#### (13) 图形更改跟踪器

图形更改跟踪器能跟踪由模型变动而引起的工程图纸的改变,并在图形上做出详细标记,同时提醒并指导设计师修正,保证图纸与模型一致。

### 7.2.4 Solid Edge 焊接设计

同钣金一样, Solid Edge 也提供一套专用命令用于焊接件的设计。焊接环境在指定焊接件、焊缝、表面处理和后处理等各环节中都提供了辅助。Solid Edge 的制图模块可产生焊前和焊后视图(图附-1-11),从而完整表达焊接工序。在装配时,焊接件是以一个单独的零件来处理的。



图附-1-11

#### (1) 专业焊接设计工具

焊接设计模块使用焊接管理器对焊接进行集中控制。从焊接构件管理、焊前表面预处理、添加焊料到焊后机加工等,各命令按钮依据实际焊接工序顺序排列,点击焊接管理器上方的按钮,会激活焊接管理器中相应的工序项,同时,焊接工具条上相关的命令按钮也将激活。与加工过程一致的控制界面既便于学习,又提高了设计效率。

#### (2) 焊接构件管理

使用插入命令可以将装配体中的所有零件或部分零件插入到焊接文件中。

#### (3) 表面预处理

使用倒角、开孔、除料等命令对需行焊接的构件表面进行焊前处理。

#### (4) 添加焊料

在焊接处可以添加焊料以反映焊接后机件的真实状况,也可以对焊缝进行标注,以便在生成工程图时直接引用。

#### (5) 焊后机加工

焊接以后,可以对焊接件进行开孔等机加工,此时添加的特征不会对原来的零件产生影响。

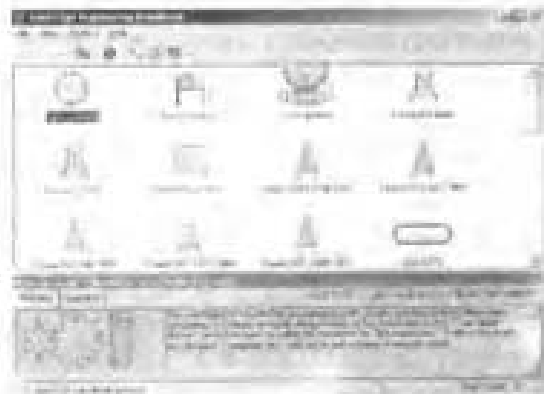
#### (6) 工程手册

工程手册是 Solid Edge 的一个附加模块,它提供在线参考,并可自动生成三维零件。工程手册由 MS 公司设计,提供了范围广泛的零件设计,包括标准公式计算和物理特性计算。根据输入的数据和设计准则,工程手册能够自动生成由计算结果驱动的三维零件。另外,工程手册中还包含了有关算法、公式和机构原理的完

整参考。

#### (7) 计算结果

Solid Edge 工程手册能对标准数学公式和有关物理算式进行计算,并依据计算结果来评估设计或改型的可行性。多数的工程计算会自动在 Solid Edge 装配图中生成参数化零件(见图附-1-12),计算结果可以保存下来,也可以插入到装配图中。



图附-1-12

工程手册提供对直梁、轴承、滚动轴承、弹簧、高速齿轮、传动带、节点、关键点、顶针点、受压节点、轴、键槽、断开键连接、受拉弹簧、焊接节点、螺旋弹簧、V 带、焊接等零件的计算。

#### (8) 零件生成器

零件生成器能依据计算结果生成参数化零件。

#### (9) 在线工程手册

在线工程手册涵盖了一系列计算公式、相关算法和理论。如果手册中不存在需要的计算公式,用户可以自己定制,也可以用 Excel 表格来控制零件或装置参数。

### 7.2.5 Solid Edge 钣金设计

Solid Edge 钣金件设计环境使用了专业的钣金和冶金术语,提供了流线性造型命令,如平板、折弯、气圆、压延、冲料、斜角、角切除和其他钣金特征等。通过自动添加弯曲形变、弯曲计算和展开, Solid Edge 提供了最先进的钣金 CAD 工具。

#### (1) 全面、专业的钣金设计命令

钣金模块提供了一个专业、高效的钣金设计环境,可以容易地进行各种钣金件设计。Solid Edge 提供的钣金设计命令是迄今为止最具效率的一整套钣金设计工具,能大大减少设计时间,提高生产效率。

从简单的平板建立到添加各种折弯、除料、倒角、开孔;从建立各种冲压特征直到自动展开,生成工程图,各种命令一应俱全。Solid Edge 钣金模块是完整的从钣金设计到制造的解决方案。

### (2) 自动处理弯曲形变

在建立折弯时,能够自动处理弯曲形变,并可加入折弯切口。通过折弯属性对话框,可以指定折弯半径和折弯切口的形状。

### (3) 斜角接缝

在建立连续折弯时,智能单步斜角接缝选项简化了复杂接角的建立。

### (4) 简洁的多重折弯

使用多重折弯命令可以一次建立多个连续折弯,二次折弯命令可以添加两个 90°折弯。

### (5) 接角缝合

接角缝合命令能使在接角处接触的折弯缝合起来。

### (6) 钣金除料

在钣金件上建立打孔、开槽等除料特征十分容易。如果除料特征跨越折弯特征,只要使用展平折弯命令先将折弯展平,然后除料,最后再使用回折命令恢复到原来的折弯状态就可以了。

### (7) 冲压特征

钣金模块提供了建立气窗、饰条和冲压特征的造型命令,操作简单,功能强大。

### (8) 钣金展开

钣金件的三维模型及其展开可保存在同一个钣金文件中,借助特征管理器能方便地在二者之间转换。在生成工程图时,系统会自动辨别钣金文件是否具有折叠和展开两种模型,并提示用户做出选择。钣金展开图与设计钣金件模型之间相互关联,若模型发生变动,展开将自动更新。

### (9) 全方位的钣金设计

钣金模块不但可以建立钣金构件、生成用于制造的展开图和工程图,而且通过与其他应用程序集成,可以实现计算分析和数控加工等功能,为钣金设计提供全方位的支持,使其降低费用、提高质量、缩短开发周期。

## 7.2.6 Solid Edge 网络发布者

网络发布器是 Solid Edge 的一个附加模块,使用人员无需具备任何网络发布的专业知识,就能快速生成含有三维模型及信息的网页。网络发布器使设计信息在互联网、公司内部网或网络服务器上的交流变得极为快捷和简单。

### (1) 网络发布向导

网络发布向导是生成含有 Solid Edge 三维模型网页的一个综合工具。在向导的指引下,只要键入必要的用户信息,即可轻松地建立网页。用户可以在向导内预览并选择预先定义的模块来构成其网页,或者以生成

自己定义的格式生成网页(图附.1-13)。

### (2) 小数据量的三维文件格式

Solid Edge 以小数据量的三维文件格式发表模型以增加查看和操作能力。网页中附有观察模型的命令按钮,可以对网页中三维模型进行动态观察。

### (3) 网页浏览

使用微软的网络浏览器 IE4.0 或更高版本可以浏览网络发布器上发布的网页,但查看 Solid Edge 设计数据需要一个免费的查看器——IPA Webview。当首次浏览网页时,浏览器会自动从 Immersive 设计公司的网站 [www.immersivedesign.com](http://www.immersivedesign.com) 上下载该查看器。查看器驻留在浏览器中,为用户提供了一个互动的观察窗口。

### (4) 产品设计交流

网络发布器的网页允许任何具有访问权的用户查看三维模型、产品结构、物料清单和设计文件属性。互动的查看视窗提供了产品的树状视图和物料清单,便于用户了解装配结构信息和产品的零件构成。

### (5) 三维模型的观察

在网页中,用户可以使用多种命令对三维模型进行控制,如放缩、旋转等,还可以有选择地隐藏或显示零件,以便于查看内部的结构。查看器还具有重新放置装配零件、生成分解图等功能。

### (6) 设计与制造同步的应用软件

网络发布器使任何拥有网络浏览器的人都可以通过互联网访问三维设计数据。它支持分发信息,分发对象包括整个企业内的非工程师、销售人员、客户及供应商。网络发布器能通过 VB 编程扩充,成为在线销售配置和其他通过互联网使用三维模型数据的协作或商用软件。

## 7.2.7 Solid Edge CAD 集成

Solid Edge 提供了一个全新的数据集成平台,可以方便迅捷地实现 Solid Edge 与办公自动化软件和其他 CAD、CAM、CAE 软件的集成。

### (1) 办公自动化软件

具有在工程图中插入 Microsoft Word 文件,在零件造型中插入电子表格,使用邮件发送程序共享设计结果,以及将设计结果以网页的形式在网上发布等功能。

### (2) CAD、CAM 和 CAE 软件

Solid Edge 采用 Parasolid 作为造型核心,可以与所有基于 Parasolid 的 CAD、CAM 和 CAE 软件集成。

### (3) 数据转换

Solid Edge 内置的双向数据转换工具使得设计团队能方便地同其他 CAD 系统交换数据,其中包括了



广泛使用的 CAD 数据格式, 如 AutoCAD (DXF/DWG), ProENGINEER, IGES 和 STEP 中性数据等。

另外还提供了 Parasolid 与 ACIS 的转换工具, 能够与基于 ACIS 造型软件的数据实现共享。



图附-1-13

#### (4) 特征识别

附加的特征识别模块具有将三维设计数据转换为基于特征的参数化设计的功能。特征识别可以自动进行, 也可以交互进行。特征识别模块使得设计师能够方便地利用以往的设计成果, 同时也可以使在不同造型系统之间的数据转换更加流畅。

#### (5) Voyager 计划

Voyager 计划将 Solid Edge 与其他软件和硬件制造商联系在一起, 为 Solid Edge 客户提供优秀的无缝集成方案。Voyager 成员提供的方案玩林广泛, 包括: 有限元分析 (FEA)、产品数据管理 PDM/ERP、设计工具、模具设计、动画制作、运动学动力学分析、逆向工程、标准件、计算机辅助制造 (CAM)、运动学转换器、机械设计、PDM、浏览/红线/动画、硬件解决方案、快速成型等。

### 7.2.8 Solid Edge Design Space

Design Space 是 Ansys 公司发布的 Solid Edge 无缝连接的产品仿真软件。使用 Design Space, 就能在设计初期, 对产品进行仿真测试, 减少了样机试制与测试



图附-1-14

的麻烦及相关的费用支出。

#### (1) 强大的仿真功能

Design Space 容易使用且功能强大, 它提供了一个独特的虚拟实验环境, 对产品在线性载荷、热、模态、疲劳和形状优化作用下的性能进行仿真 (图附-1-14)。

#### (2) 用户界面友好

Design Space 完全融入在 Solid Edge 环境中, 通过基于知识的自动化技术和互动向导的结合, 使得

Design Space 的操作十分简单。在向导循序渐进的指引下,可以轻松实现整个仿真过程。在仿真过程中,可以随时保存纪录、图片或文档,因而可以创造一个更加程式化的、完全归纳的全自动仿真过程(图附-1-15)。



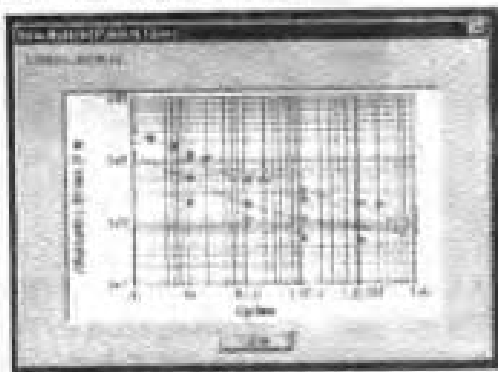
图附-1-15

#### 智能集成

Design Space 提供了 CAD 工具与仿真模块的完全继承,能够快速、清楚地仿真设计的修改。Design Space 能记忆加载点和支撑点的位置,这样在变更设计后,无需再次重复仿真设定过程,只要点击鼠标,就能得到新的结果。

#### (3) 双向关联

由于采用了参数化仿真功能,Design Space 和 Solid Edge 能够双向关联。使用 Design Space 的参数管理表(图附-1-16),用户可以输入多种设计规则,只要点击一下鼠标,就能生成新的模型并进行分析,这对于评估多种设计方案极为有利。



图附-1-16

#### (4) 模型转换

如果需要非线性有限元分析,无论处在何种阶段,Design Space 都能智能地将当前的成果转换为

象 ANSYS 一样的高级 FEA 软件。

#### (5) Solid Edge Insight——与 CAD 集成的 PDM 新工具

Solid Edge Insight 是将产品数据管理技术融入到设计师日常 CAD 工作中的一项新技术。这是一个革命性的工具,使 CAD 工程师能在不知不觉中轻松实现 PDM 功能。

与通常所见的 PDM 系统不同,Solid Edge Insight 不是一个独立的软件包,使用它无需增加许多命令和费时的实现过程,恰恰相反,Insight 只是集成在 Solid Edge 中的一个模块,不需要额外付费就可实现,它也无需用户培训,可在数小时而不是数月或数年就能完成。Insight 的主要功能包括:

##### 1) 为设计师服务的数据管理

Solid Edge Insight 用一种新的方式管理产品数据,使设计师能够集中精力关注工程设计。

##### 2) 全面提高设计效率

Insight 的引入,使得 Solid Edge 已不再是单纯的 CAD 系统,它提高了整个设计团队的效率。通过搜集和管理设计单位的综合数据,Solid Edge 提供了加速装配设计的关键工具。

##### 3) 快捷简单的实施方案

根据工业分析,大多数产品数据管理需要一年或更长时间才能实现,而 Insight,则是一个简单快速而且易于实现的产品数据解决方案。

##### 4) 额外的投资回报

同当前 CAD 或 PDM 系统相比,Solid Edge 廉价的软件投资、实施、培训和系统管理费用,使用户能获得更丰厚的回报。

#### 7.2.9 Solid Edge Gibbs CAM

Gibbs CAM 是易学、易用、易懂的计算机辅助虚拟加工系统,它采用 Parasolid 实体造型核心,可与 Solid Edge 无缝集成。Gibbs CAM 具有独具匠心的图形化界面,即使没有电脑基础的人,亦能驾轻就熟地使用 Gibbs CAM 进行加工。Gibbs CAM 具有以下特点:

##### (1) 图形化界面与零件加工向导

填空式的设计选择,使您只要看图填空,就能轻松操作。

##### (2) 图形化刀具库及复式加工

全图形化的弹性刀具库设计只要看图选取图形即可加工。另外还可自由设计成形刀及自由组合设计加工方式并存在加工文档供下次使用。

##### (3) 超强读档能力

Gibbs CAM 具有运算 24 次 Nurbs 高阶曲面的能力,可完整读出高档 CAD 系统输出的 IGES 档案。所

看 Parasolid 及 ACIS 核心版本都可直接读取。

#### (4) 快速缝补曲面及实体补洞功能

利于实体分模能有效地将任意相邻曲面加以缝补,可以避免不必要的掉刀问题,更能增加加工的面滑度,节省大量的路径运算时间,更可利用实体补洞功能及布尔运算快速分模。

#### (5) 全功能的 2D~2.5D 切削

Gibbs CAM 可快速建立 2D~2.5D 刀具路径,包含各式钻孔、挖槽、轮廓、铣削,自动拉出粗切路径,自动留出岛屿深度的(岛屿数目不限)成形刀、铣刀、倒圆角及斜面之断面形状,沿曲线做 2.5D 扫描之路径等路径功能(图附-1-17)。

#### (6) 全功能的 3D 曲面及实体切削

Gibbs CAM 可切削曲面模型,更可直接切削实体

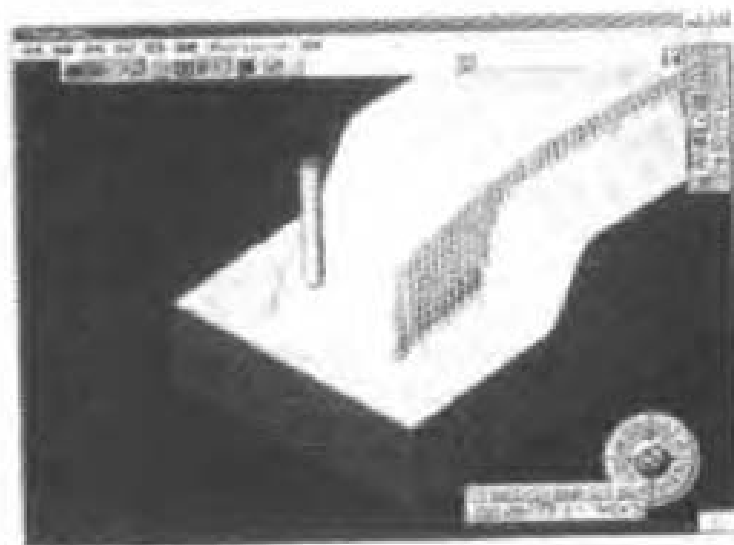
模型(图附-1-18),环绕式等Z及不等Z降层粗切,投影式等Z及不等Z降层粗切,插削粗切,粗切预先自动钻孔与下刀点,下刀有弧向直线下刀,链接式斜向进刀,螺旋进刀及路径最佳化功能以增减切削时间,精切法有等距投影,不等距粗糙度控制投影,等高降层及不等高降层粗糙度控制加工投影,指定范围加工,两曲线沿旁式或穿路式刀具路径,放射状及环绕式路径,曲面膜向路径,指定清角,投影式残料加工,粗切式残料加工等。

#### (7) 支持多种进刀法及最佳化路径

Gibbs CAM 拥有最新的下刀法,可有效地快速下刀,更聪明的是,能自动找出最省时间的下刀位置,而且会再下刀点自动找出并钻出不同的下刀孔(图附-1-19)。



图附-1-17



图附-1-18



图例 1-19

#### (8) 支持 NURBS 高速曲线切削功能

Gibbs CAM 是 Fanuc, Mitsubishi, Siemens 三大控制器指定的内建于新一代 PC-Based 控制器上的 CAM 系统,由核心直接运算出 Nurbs 曲线路径,比由直线段换成曲线的路径更快、更省、更精确。

#### (9) 人工智能 NC 程序与加工计划

人性化的加工参数化设计,使您可以随时依加工条件的不同更新刀具路径,更可以随时变换刀具或是补正量,只需变更一下刀具直径或是补正量,轻按一下重新计算按钮,程序就会自动更新,基于机件的刀具路径,无论是图形改变、三维模型改变,刀具改变,素材改变都可全部重新计算。

#### (10) 超强的 3D 实体切削模拟

真正的 3D 实体切削模拟,不需要出 NC 程序即可模拟,还可以快速发觉加工问题,大量降低加工的失败率,达到真正的效率优化。

#### (11) 后处理支持的格式

支持车床、三轴铣床、两轴定位加工铣床、四轴联动加工铣床、五轴铣床及车铣复合机种等程序输出,提供 3700 种以上不同品牌的机种进行个别的处理。

### 7.2.10 Solid Edge Dynamic Designer

Dynamic Designer/Motion 是美国 MDI 公司 (Mechanical Dynamic, Inc) 为中档 CAD 市场重点推出的全功能三维机构动态仿真软件。它的内核为世界上最具权威、使用最广的 ADAMS (MDI 公司的高档产品) 动态求解器,功能强大、仿真可靠。通过与当今主流三维 CAD 软件无缝集成,用户无需离开自己熟悉的 CAD 环境就能使用它,操作简单快捷。

使用 Dynamic Designer,在设计变成产品之前,您可以看见并研究装配体在运动时的情况,诸如零件之间是否存在运动干涉,承受的载荷多少之类的问题就迎刃而解了。

Dynamic Designer 产品完全集成于 Solid Edge 环

境,它包括 Simply Motion, Motion 和 Professional,其中 Simply Motion 已存在于 Solid Edge 之中。随着设计复杂程度的增加,或需要了解更多的产品功能信息,可以逐步升级到更高一级的产品。

Simply Motion 可以仿真各种运动联接,发动机、激励和弹簧等系统的实际工作情况,更容易地让您的设计模型运动起来,快速地检查零部件之间的干涉,它实用的三维动力学可实现比仅仅是运动学工具更广泛范围的装配体动态仿真。

Dynamic Designer/Motion 增强了 Simply Motion 的功能,可以研究更复杂的系统,例如,含凸轮、齿轮、拨叉机构和微摆等,它是全功能的三维动态仿真工具,能仿真零部件间力和碰撞的影响,并将仿真结果通过多种绘图方式显示出来。

Dynamic Designer/Professional 能得到有关设计产品性能的更多的参数信息,通过大小,得到消耗能量的多少来生成用于有限元分析的工作载荷。

#### Dynamic Designer 产品的特点

1) Solid Edge 认证的首选合作伙伴软件可与 Solid Edge 集成为单一界面。

2) 丰富的产品线——Simply Motion, Motion 和 Motion Professional 允许您选择能满足您现在需求的产品,并确信当您的需求进一步提高时还有更高级的产品能满足它。

3) 丰富的运动定义可以解决各种工程问题,从简单机构到包含齿轮、凸轮、拨叉和各种碰撞的复杂装配体。

4) 自动从 Solid Edge 装配体直接映射生成功能仿真模型。

5) 多种有效的用于物体间碰撞建模的碰撞定义,适用于点对点的碰撞的简单冲击力,适用于平面内碰撞的 2D 曲线碰撞,还有通用的 3D 表面碰撞。

6) 所有的约束和碰撞都支持摩擦定义。

7) 材料数据库包括常用材料的碰撞和摩擦参数。

8) 通用 ADAMS 求解器,由此提供并保证可靠、准确而有效的动力学分析解算。

9) 求解的结果包括位移、速度、加速度和反作用力。

10) 生成设计产品的运动动画,并保存为 AVI 格式或便于通过网络协同合作的 VRML 格式。

11) 检查设计的产品在工作流程内各零部件间的运动干涉。

12) 为有限元分析提供准确的载荷。

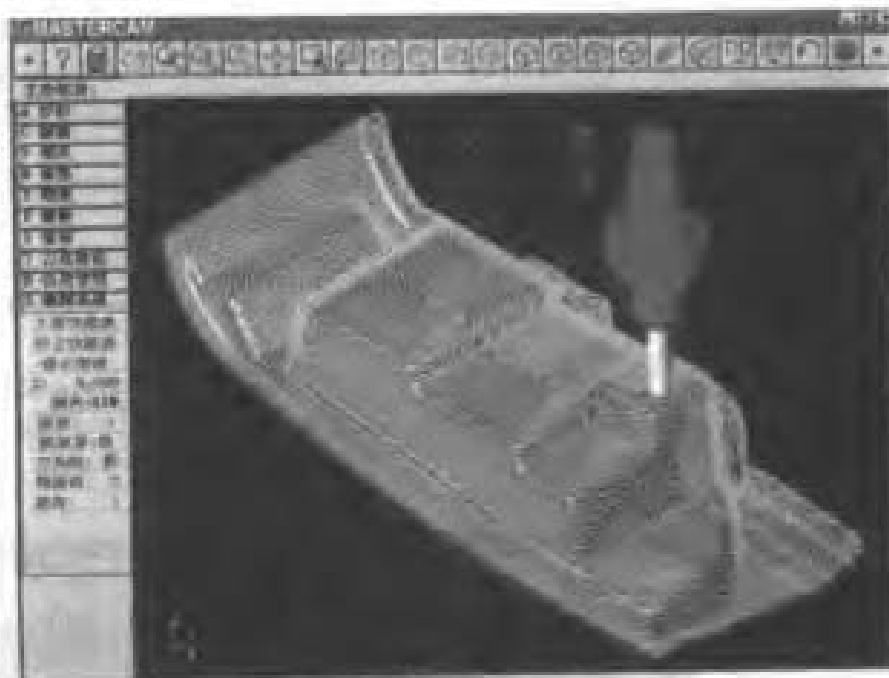
## 8 MasterCAM 软件

MasterCAM 是美国 CNC 系统公司开发的一套适用于机械产品设计、制造的运行在 PC 平台上的 3D

CAD/CAM 交互式图形集成系统(图附-1-20)。它不仅完成产品的设计,更能完成各种类型数控机床的自动编程,包括数控铣床(2~3轴)、车床(可带C轴)、线切割机(1轴)、激光切割机、加工中心等的编程加工。

产品零件的造型可以由系统本身的CAD模块来建立模型,也可通过三坐标测量仪测得的数据建模,系

统提供的 DXF、IGES、CADL、VDA、STL 等标准图形接口可实现与其他 CAD 系统的双向图形传输,也可通过专用 DWG 图形接口直接与 AUTOCAD 进行图形传输。系统具有很强的加工能力,可实现多曲面连续加工、毛坯粗加工、刀具干涉检查与消除、实体加工模拟、DNC 连续加工以及开放式的通用后置处理功能。



图附-1-20 MasterCAM 的加工仿真

## 8.1 CAD 设计功能

- 1) 2D、3D 设计及尺寸标注,输出工程图。
- 2) 可构建 3D 线框造型,曲面造型和实体造型。
- 3) 曲面造型可生成直纹曲面、举升曲面、扫描曲面、旋转曲面、昆氏曲面、牵引曲面以及对曲面修整、熔接、投影、修正和多曲面倒圆角。
- 4) 具有非均匀 B 样条 NURBS 曲面的构造能力,以满足复杂曲面的构造及制造需求。
- 5) 实体造型可对实体布尔运算、倒圆角、延伸和修整。
- 6) 可自定义方程式,构建复杂的 2D、3D 模型。
- 7) 有动态导航功能,可动态捕捉特殊点。
- 8) 具有多种过滤方式(图层、颜色、类型)来选择因素。
- 9) 可对图形作缩放、旋转、缩放、平移等转换处理。
- 10) 可构建 True type 的所有字符。
- 11) 具有与 C 语言的二次开发接口。

(2) 对 3D 图形可作渲染、着色处理。

## 8.2 CAM 加工功能

- 1) 2D、3D 外形铣削加工包含刀具半径补偿和转角计算。
- 2) 钻、攻、镗等固定循环加工。
- 3) 带多高的往复式及环绕式挖槽加工,可螺旋式下刀,外轮廓和岛屿可设定倾斜角加工。
- 4) 单曲面、多曲面粗加工和精加工。
- 5) 提供环绕等距加工,平行式加工,放射状加工,等高加工等加工方式。
- 6) 提供铣削及清角加工,圆拉刀方式加工,投影加工,沿面加工,挖槽式加工,浅平面及倾斜面加工等加工方式。
- 7) 支持高速加工方式。
- 8) 曲面加工过切检查及消除。
- 9) 建立具有关联性的操作管理,集刀具、工件材料、加工参数、刀具路径和串联的几何图案于一操作组,任一参数修改不必从头做起,只要更新操作就可生

成新的刀具路径。

10) 用户定义加工坐标系及选择刀具平面。

11) 修改、定义刀具库和材料库,可自动生成推荐的主轴转速及进给率。

12) 动态刀具路径模拟。

13) 刀位中间文件和数控加工文件的双向转换。

14) 编辑刀位文件并对刀具路径作镜像、旋转、缩放、平移等转换处理。

15) 在给定的误差范围内对刀位文件作过滤处理,可缩短加工程序。

16) 提供 400 多种专用后处理程式及开放式的通用后处理程式,可适用于各种型号的数控系统。

17) 产生加工参数报表,计算加工时间,可供操作者参考。

18) 实现计算机与数控机床的双向数据传输及 DNC 边加工边传输。

## 9 EdgeCAM 数控自动编程系统

EdgeCAM 为英国 Pathtrace 公司出品的数控自动编程系统。EdgeCAM for MDT 集成于 Mechanical

Desktop (MDT), 直接对实体模型进行编程, 自动识别特征, 极大地提高了生产效率。

### 9.1 EdgeCAM 软件特点

1) 在 MDT 环境中, 直接进行数控自动编程, 无需转换界面, 无需数据转换, 易学易用。

2) 智能特征识别。自动从实体获取加工信息, 自动编程, 大大减少了人工操作的错误。

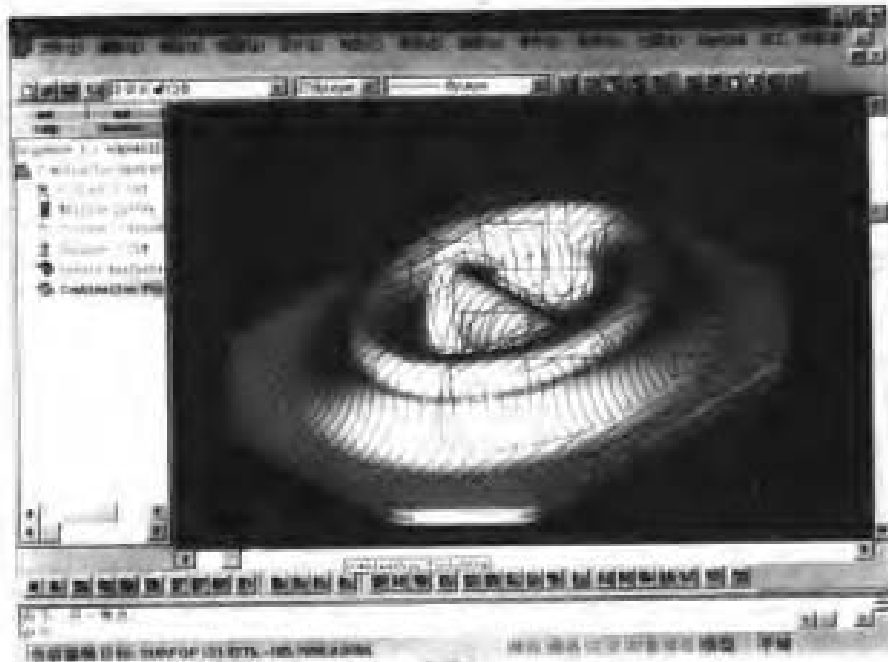
3) 加工方法自动施加于特征, 大大减少人工操作的时间, 例如对一个较复杂的孔的编程, 常规操作需要上 20 条指令, 而在 EdgeCAM 内只要一项操作即可完成。

4) 刀具路径与实体模型动态关联。若修改实体模型, 刀具路径自动更新。

5) 对于用参数化设计的系列零件, 只需一次编程。

6) 确保数据的完整。刀具路径与设计模型共存于同一个文件中。

7) 设计与制造联系更紧密。EdgeCAM 支持并行工程, 在设计早期就可决定加工方法及编程, 避免错误的发生, 缩短产品上市时间。



图附-1-21 EdgeCAM 加工的零件

8) 支持当今最通用的各种加工方法 (图附-1-21), 2 轴车; 4 轴车, C/Y 轴车; 2、5 和 3 轴铣; 曲面加工; 旋转铣, 多面加工; 5 轴铣; 2-4 轴线切割。

### 9.2 系统功能

(1) EdgeCAM 基本特性

- 1) 支持半径补偿, 固定循环加工, 做套子程序。
- 2) 支持数控机床与 PC 的通信。
- 3) 简单快速地编辑加工指令。
- 4) 集成的加工浏览器, 使得加工操作的编辑、重排序更加简单。
- 5) 包含 NC 代码编辑器。

## (2) 2.5 和 3 轴铣

1) 单独记录每次操作, 编辑及重新排序非常容易。

2) 使用简单智能的操作产生 NC 程序。

3) 轮廓加工具有刀具偏置计算, 精加工公差, 自动切削补偿模拟移动, 倾斜引入引出线的选项。

4) 在粗加工中完全支持内岛, 减少进刀与退刀时间。

5) 智能型腔加工, 带有最大步长自动型腔加工, 可减少 33% 加工时间。

## (3) 多面加工

1) 支持零件不同面上的孔、型腔和曲面加工。

2) 减少装卡, 充分发挥机床所有功能。

## (4) 曲面加工

1) 曲面加工具有多面过切探测能力。

2) 多种刀具切入方法改善了加工质量延长刀具寿命。

3) 多种曲面精度控制方法, 包括流线和水线加工在内的多种加工策略。

4) 毛刺高度控制保证了曲面稳定的光洁度。

5) 等高截面加工可快速清除材料。

6) 完全避免过切。

7) 支持各种刀具。

## (5) 高级曲面加工

1) 智能组合精加工——在同一路径中智能自动组合不同的加工策略, 减少加工时间, 提高加工质量, 减低使用者技术要求。

2) 专门为模具业设计的高级加工策略。

3) 高级投影加工循环——优化加工策略, 提高加工效率。

4) 等高凹模和凸模粗加工。

5) 自动残余量加工——自动检测未切削区域, 减少精加工时间, 简化编程操作。

## (6) 车削

1) 支持所有的 2 轴车床所支持的加工方法。

2) 起刀点与退刀点可在沿着轮廓的任何地方。

3) 自动辨别插入的几何体, 消除了过切的可能性。

## (7) 2 轴车

1) 在刀具路径计算中考虑整个刀具形状。

2) 避免过切。

3) 自动毛坯更新减少空切时间。

4) 端面加工, 回转面加工, 槽加工大大减少用户的输入。

5) 以工业标准定义刀具。

## (8) C/Y 轴 (驱动工具)

1) 使绕直径的 3D 旋转铣编程更加容易, 这是通过让 2D 图形绕直径旋转而得到。

2) C/Y 的运动可在轴向和径向编程, 可使用所有的铣、孔加工方法。

## (9) 4 轴车

1) 平衡镜像切削适用于大容量的切削。

2) 双刀架同步。

3) 支持同轨双刀架。

## (10) 旋转加工

1) 绕圆柱零件加工孔与型腔。

2) 支持旋转工作台。

3) 完全的旋转刀具路径模拟。

4) 自动跟踪基准点与原点偏移极大的减少编程时间。

## (11) 螺纹铣

1) 具有内、外和单/多头螺纹加工能力并完全支持固定循环。

## (12) 线切割

1) 为切割零件提供了标准的格式, 大大缩短建模时间。

2) 2~4 轴线切割。

3) 最少的设置参数大大简化数据输入。

4) 支持圆锥切割, 支持上下轮廓形状不同的切割。

5) 支持截面切割, 支持断前留量切割和捣毁切割。

## (13) 刀具库

1) 具有直观图形化的界面, 可方便地保存、更改、扩充和调用刀具。

2) 支持铣削 (含孔加工) 和车削刀具。

3) 可以保存切削参数赋予刀具。

## 9.3 EdgeCAM 实用工具

## (1) 代码向导 (Code Wizard)

使用数控代码向导, 你可使用简单的模板为几乎所有的 CNC 机床器配置数控代码生成器 (后处理器)。

## (2) 数控代码生成器

1) 专为 NC 代码配置的编辑器。

2) 不同颜色的文字更易于浏览。

3) 可重新对数控代码进行排序、修改。

## 10 CAXA 系列 CAD/CAM/CAE 软件

## 10.1 CAXA 概要

CAXA 电子图板和 CAXA-ME 制造工程师软件的开发与销售单位是北京北航海尔软件有限公司 (原

北京航空航天大学华正软件研究所)。

CAXA 电子图板是一套高效、方便、智能化的通用中文设计绘图软件,可帮助设计人员进行零件图、装配图、工艺图表、平面包装的设计,适合所有需要二维绘图的场合,使设计人员可以把精力集中在设计构思上,彻底甩掉图板,满足现代企业快速设计、绘图、信息电子化的要求。

CAXA-ME 是面向机械制造业的自主开发的、中文界面、三维复杂形面 CAD/CAM 软件。

## 10.2 产品设计

CAXA 电子图板 2000 是适用于通用绘图和设计的 CAD 软件,可用于机械制图、电气制图、建筑制图、平面包装等所有需要二维绘图的场合,是一种高效、方便、智能化的设计和绘图工具。CAXA 电子图板 2000 是 CAXA 电子图板的最新版本。

强大的三维产品设计功能。CAXA 提供强大的三维 NURBS 曲面造型。从曲线到曲面,提供了丰富的建模手段。可通过列表、数据、数学模型、字体文件及各种测量数据生成样条曲线;通过扫描、放样、拉伸、导动、等距、边界网格等多种形式生成复杂曲面;并可对曲面进行任意裁剪、过渡、拉伸、变形、相交、拼接等,建立任意复杂的零件模型。通过曲面模型生成的真实感图,可直观显示设计结果。

功能强大的项目管理。CAXA 电子图板 2000 提供了强大的图样和项目管理功能,快速检索工具可按图样名称、设计人、图样代号、幅面、日期等指定条件组合检索,查找图形更快捷。图样可设置口令,有效保护企业机密。自动生成数据库,对项目进行系统的管理。标题栏和明细表与数据库关联,可以根据产品部件图明细表,自动生成总装图明细表。对产品中的标准件和非标准件,可进行统计并自动生成统计报表。

全面开放的开发平台。CAXA 电子图板 2000 充分考虑用户的个性化需求,提供了专业而易于使用的二次开发平台,以实现产品的个性化和专业化。您可以随心所欲地扩展电子图板的功能。甚至,在 CAXA 电子图板 2000 强大的二次开发平台下,您可以开发出全新的 CAD/CAM/CAE 产品。

建筑电气的专业模块。CAXA 电子图板 2000 扩充了专业的齿轮、建筑和电气设计模块,除原有的 16 大类万余个规格系列的参量化国标机械和电气零件图库,又新增了电机、管接头、机床夹具、变压器、触点、二进制逻辑单元、模拟单元等数千个规格的机械和电气常用标准件。这些模块和图库特别适合用户进行建筑平面图、施工图和电气原理图的绘制。

DXF 和 IGES 图形接口。通过 DXF 和 IGES 标准

图形接口,CAXA 制造工程师可以与任何 CAD/CAM 系统交换数据。

## 10.3 制造工程师

面向加工中心的 CAM 软件。强大的造型功能可快速建立复杂的三维模型,具有灵活多样的加工方式,快速生成加工轨迹。通用后置处理适用于各种数控系统。支持三坐标测量机,具有数控代码反读和真实感仿真功能。CAXA 制造工程师是一个功能强大,易学易用的全中文三维复杂型面加工 CAD/CAM 软件。已广泛应用于塑模、锻模、汽车覆盖件拉伸模、压铸模等复杂模具的生产以及汽车、电子、兵器、航空航天等行业的精密零件加工。它可以在保证最好加工质量的前提下,最充分地发挥数控机床效率,缩短产品设计、制造周期。

CAXA 数控铣是面向数控铣床的自动编程系统。可对包含任意多个干涉岛的复杂形状的二维区域进行铣削加工编程,加工时可以选择单向、往复、环切等走刀方式。并可生成加工腔槽带有拔模斜度的二轴半的刀具轨迹。CAXA 数控铣系统充分利用数控机床的“固定循环”功能,实现钻孔、深孔钻、攻螺纹、镗孔等十几种孔加工方式。

CAXA 线切割是一个面向线切割机床数控编程的软件系统。CAXA 线切割为您提供线切割机床的自动编程工具提高效率。CAXA 线切割可以交互方式绘制需切割的图形,生成带有复杂形状轮廓的两轴线切割加工轨迹。CAXA 线切割支持快走丝线切割机床,可输出 3B 后置格式。输出 3D 或 G 代码,支持快、慢走丝线切割机床、水切割、火焰切割、激光切割等数控机床的编程加工。

快速高效的数控加工能力。CAXA 制造工程师快速高效的加工能力涵盖了从两轴到五轴的数控铣床功能以及数控车床、钻床、线切割和加工中心。

1) 两轴及两轴半加工方式:可直接利用零件的轮廓曲线生成加工轨迹指令,而无需建立其三维曲面模型;提供轮廓加工和区域加工功能,加工区域内允许有任意形状和数量的跳岛。可分别指定加工轮廓和岛的拔模斜度,自动进行分层加工。

2) 三轴加工方式:多样化的加工方式可以安排从粗加工、半精加工到精加工的加工工艺路线,提供粗加工、参数线加工、限制线加工、复杂曲线加工、曲面区域加工和曲面轮廓加工等多种加工方法,针对复杂模型快速、自动生成精确、无干涉的加工轨迹指令。

3) 四到五轴加工方式:针对叶轮、叶片类零件提供的多轴加工功能,可以利用刀具侧刃和刀具端刃加工整体叶轮和大型叶片。支持带有锥度的刀具进行加



工,任意控制刀轴方向。CAXA 制造工程师的每种加工方式所得轨迹代码均在实际加工过程中得到验证,每一加工功能都是众多现场编程人员和机床操作人员丰富经验的概括和总结。

控制加工工艺性。CAXA 制造工程师提供了丰富的工艺控制参数,可以方便地控制加工过程,使编程人员的经验得到充分的体现。丰富的刀具轨迹编辑功能可以控制切削方向以及轨迹形状的任意细节;可以指定垂直、直线相切、圆弧相切等各种进刀方式;可以选择单向、往复、环切、平行等切削方式;可以任意控制顺、逆铣方式,以适应不同的要求,得到高品质的加工效果。

代码校核和真实感仿真。CAXA 制造工程师提供了多种刀具轨迹仿真手段以检查数控代码的正确性。可以通过线框动态和静态仿真检验刀具轨迹在某些关键点上的切削状态。也可通过实体真实感仿真如实地模拟加工过程,展示加工零件的任意截面,显示可能出现的干涉现象,从而减少在时间、资金和材料的浪费。对于已有的数控代码,也可通过 CAXA 制造工程师的代码反读功能将数控代码转换成图形显示的加工轨迹,真实仿真切削结果。

通用后置处理。CAXA 制造工程师开放的后置处理使用户可以配置针对任意机床的后置处理程序,以适应各种不同的数控系统。CAXA 制造工程师充分利用数控机床的固定循环功能,生成高效、优化的数控代码,提高复杂加工的代码效率。

注塑工艺设计是与美国 C-MOLD 公司合作开发的面向注塑行业的 CAE 软件。采用国际 CAE 技术的最新成果,通过科学的分析方法和简便的操作,可预测注塑工艺过程,确定优化的注塑工艺参数,达到优化设计的目的。

CAXA 注塑模设计是一套中文注塑模专业 CAD 软件,该软件提供注塑模标准模架和零件库,以及塑料、模具材料和注射机等设计参数数据库,可随时查询、检索;并能自动换算型腔尺寸,对模具进行各种计算。使用该软件,设计人员不必翻找设计手册即可轻松设计模具。

支持测量数据处理。CAXA 制造工程师可读取三坐标测量机和扫描仪获得的数据,对数据进行平整和光滑处理,并进行蒙面,再现被测产品外形,取得满意的测量效果。

#### 10.4 其他功能

丰富流行的数据接口。CAXA 电子图板 2000 是一个开放的设计/绘图工具,提供了丰富的数据接口。它们包括了传统的 DXF 和 DWG 接口,面向工作站

CAD/CAM 软件的 IGES 接口,面向办公和大量文字处理的 TXT 接口和面向大量数据和图纸管理的 DBF、MDB 接口。这些接口保证了与世界流行的 CAD 软件进行双向数据交换,使企业可以跨平台和跨地域地与合作伙伴实现虚拟产品开发和设计。

低成本的运行环境。CAXA 制造工程师以 PC 微机为硬件平台,适合国情特点 最低要求:80486DX 处理器,8MB 内存;40MB 硬盘;VGA 彩显;两键鼠标。推荐配置: Pentium200 处理器,64MB 内存;1GB 硬盘。CAXA 制造工程师可运行于 DOS、WINDOWS、WIN95 和 NT 等系统平台之上。

易学易用的全中文软件。CAXA 制造工程师采用图标、菜单及键盘结合的交互方式,功能紧凑、方式简洁。既考虑了初学者的需要,又为熟练者提供方便。CAXA 制造工程师是从底层开发的全中文 CAD/CAM 软件,无需依赖汉字操作系统也无任何汉化工作,工作稳定可靠。系统菜单、状态、提示信息、全程在线帮助及使用指南均为全中文表达,适合国情,使工厂的数控机床编程和操作人员能够很快掌握。

## 11 金银花 CAD/CAM 系统 LONICERA

### 11.1 金银花 LONICERA 软件概述

金银花 LONICERA 系列软件是基于 STEP 标准的 CAD/CAM 集成系统,是国家 863/CIMS 计划“九五”期间最新研究成果,是面向 20 世纪的支持企业信息集成的机械产品设计与制造一体化软件。

该软件主要用于机械产品从设计到制造全过程中,实现设计/制造一体化及自动化。该软件技术起点高,以制造业最新国际标准 ISO10303 (STEP) 为系统设计的依据;采用面向对象的技术,使用行业内先进的实体建模、参数化特征造型、二维与三维一体化、SDAI 标准数据存取接口等技术;具备机械产品设计、工艺规程设计和数控加工程序自动生成;同时还具有多种标准数据接口 STEP、DXF、SAT 等,支持产品数据管理 (PDM)。目前金银花系统的系列产品包括:机械设计平台 MDA、数控编程系统 NCP、产品数据管理系统 PDS、工艺设计工具 MPP。

系统特点有:

1) 符合产品数据表达与交换国际标准 ISO10303 (STEP),遵循 AP214/AP224/AP203。

2) 采用“面向对象”技术,应用“软件组件”概念,开放性好。

3) 产品设计、工艺规程设计、数控加工编程和有限元分析等应用软件组件插入金银花系统框架即可方

便地实现产品数据的共享和集成。可支持企业开发产品的并行工程。

4) 用户界面形象化, 具有设计过程导航操纵器 (DEP), 易于学习、使用方便。

5) BEE-DAE 模块按照 STEP 的标准数据存取接口规范 (SDAI) 技术, 可完成对全系统数据的存取和管理。

6) 机械设计平台 (MDA-CADE) 具有三维实体建模、参数化特征造型、变量化设计并具备三维/二维双向关联、自动转换绘图。

7) 具有 STEP 中性文件和各种数据转换接口, 可与其他流行的 CAD/CAM 软件 (如 Pro/ENGINEER、UG-II) 进行数据交换。

## 11.2 金银花 LONICERA 软件功能

机械设计平台 MDA (Mechanical Design Assistant) 是金银花 LONICERA 系列软件之一, 是二维与三维一体化设计系统。机械设计平台 MDA 开发平台具有强大的设计功能, 主要提供如下三大模块: 零件设计、工程图设计、产品装配设计。

### 11.2.1 零件设计模块

零件设计模块由草图设计模块和特征设计模块组成。草图设计模块定义三维空间平面中二维的轮廓, 经拉伸、旋转、扫描生成三维实体 (即特征)。特征设计模块定义三维基本特征、基准特征、高级曲面特征。一般情况下, 由草图模块提供轮廓, 经拉伸、旋转等后生成基体特征, 然后在基体特征上叠加特征 (孔、凸台、倒角特征等), 构成零件。特征设计模块同时提供修改特征、删除特征、平移特征、镜像特征、阵列特征等编辑功能。

设置图元捕捉类型。系统设有一个捕捉状态的标记, 记录当前是否处于捕捉状态以及捕捉类型。系统根据每一类特征定义特点, 设置捕捉类型标记, 如创建草图时, 系统捕捉类型设置为捕捉几何平面或基准 (辅助) 平面, 其他任何类型均被过滤, 避免选错; 创建倒角、圆角特征时, 系统捕捉类型自动设置为实体包含的边, 其他任何类型均不被捕捉。创建草图特征时捕捉基准面, 创建倒角圆角特征时捕捉实体上边。

捕捉基准特征。机械设计平台 MDA 系统基准特征共分四类: 基准平面、基准轴线、基准点、基准曲线。基准平面为无界平面, 但系统以有界的矩形区域表示, 并在矩形的左下角标记其名称 (DTM) 和标识号 (如 Id=2), 构成基准面 (如 DTM2); 基准轴为无界直线, 系统以有界线段表示, 并在箭头附近标识其名称 (A-) 和标识号 (如 Id=2) 构成基准轴 (A-2); 基准点以

“X”标识号显示在屏幕上。

图元捕捉次序设定。对于一些设计状态, 如删除操作, 所有类型都处于捕捉状态, 这时系统设置捕捉优先次序。机械设计平台 MDA 系统将动态捕捉次序定义如下:

1) 草图设计模块: 点、端点、图元上点 (直线、圆弧、圆等)、图元中点 (直线、圆弧、圆等)、交点 (共面直线交点)、圆 (弧) 心。

2) 三维特征设计模块: 基准点、基准轴线、边 (直线、圆、圆弧、其它曲线、实体边界线)、基准平面、几何表面 (内外表面)、特征 (凸台、切去、孔、键槽、直腔、阵列、镜像、圆角、倒角等)。

3) 零件特征构造树: 零件构造树显示零件构造过程的结构, 树上每一节点表示一个特征, 可以从零件构造树上选择节点来选择实体中相应特征。当鼠标单击时, 该节点表示的特征即在图形显示窗口中高光显示。特征的创建及修改过程一目了然, 显示特征, 用鼠标单击特征节点, 显示特征几何形状; 显示尺寸, 用鼠标双击特征节点, 显示特征尺寸; 重命名, 可对树节点即特征重新命名; 删除, 将光标移至某一节点附近, 击鼠标左键, 在窗口内显示特征几何形状, 这时按键盘中 “Delete” 键, 系统弹出删除特征对话框, 选择 “确认” 方式后可直接对零件构造树进行删除操作。以上各项功能也可以通过鼠标的右键功能得以轻松实现。

### 11.2.2 二维工程图样

三维零件设计完成后, 机械设计平台 MDA 系统可以将其自动转化成零件的二维工程图样, 工程图样模块提供多种视图生成、编辑功能, 可以定义剖视图 (全剖视图、半剖视图、阶梯视图、局部视图、截断视图)、剖面图等, 三维尺寸经过投影生成二维尺寸, 提供尺寸编辑和视图修改功能, 生成图样符合中国国家标准, 零件设计与工程图样之间相互关联。

### 11.2.3 装配设计模块

装配设计模块完成产品从零件到部件的整个装配过程。系统定义多种装配约束关系, 装配过程中自动检查零件的约束情况; 提供的装配树使得产品结构及装配过程一目了然; 系统提供干涉检验、最小间隙、爆炸视图、剖切视图等功能及装配体的管理、编辑和修改功能。

### 11.2.4 其他功能

#### (1) 动态捕捉设计意图

机械设计平台 MDA 系统采用鼠标按键全动态捕捉方式拾取图元和各种特征。无需用多个菜单选项来

指定不同捕捉类型,也无需一个常驻的工具条,系统提供动态导航方式设计特征。

#### (2) 独特的零件设计窗口

机械设计平台 MDA 零件设计窗口分两部分:

1) 零件特征构造过程树:反映零件设计、装配及产生工程图样的层次结构表。

2) 图形显示窗口:零件的装配件及工程图样的图形显示窗口,设计过程中单个特征、尺寸亦可显示在窗口中。

#### (3) 零件、工程图样设计、装配同时设计

用户可以同时打开零件设计窗口和图样设计窗口及装配件设计窗口,不同的设计窗口对应不同的工具条(即功能菜单),只有一个窗口处于激活状态。

#### (4) 多个零件同时设计

机械设计平台 MDA 提供多文档设计功能,可同时打开多个零件、图样文件及装配件进行设计,并可用多个窗口显示出来,当多个文件同时打开时,会出现多个窗口互相覆盖问题,为使图面清晰,可从“窗口”菜单选择“水平排列”、“竖直排列”等重新排列窗口。用户可以随时改变每个窗口的大小,亦可将窗口缩小成图标显示在机械设计平台 MDA 主窗口底部。

在线中文帮助。系统提供在线联机帮助,用户有疑难问题时可随时启动该功能而获得帮助信息。

## 12 InteCAD

### 12.1 InteCAD 特点

武汉天喻公司 InteCAD 5.0 是集工程制图、参数化设计及产品信息管理三大模块于一体的二维辅助设计绘图系统。基于画法几何原理,系统功能丰富实用,涉及二维工程绘图的各个方面。面向集成。良好的开放性使用户可以任意移植、添加所需图库和模块。以 AutoCAD R13/R14 为平台。全面的动态导航,智能尺寸标注。各类工程制图方式。复杂约束模式的变量设计,强大的后参数化能力。多视图联动约束定制。基于参数化模块开发的标准件库。可网络化的零件建库及管理工具。将设计准则、过程与参数化结合 InteBasic。强大的明细表自动识别、定制填写、汇总、统计功能。与产品数据管理 IntePDM 的无缝集成。开放的设计计算工具。

### 12.2 基本功能

#### 12.2.1 全面的导航支持

设计人员在绘图过程中,往往要捕捉一系列的关键点,如端点、中点、垂点、切点等。动态导航可以在

绘图过程中猜测用户意图,自动捕捉这些关键点,不需要输入任何捕捉信息。同时导航按照长对正、高平齐、宽相等的制图规则对三视图绘制提供了十分方便的支持,用户在计算机上绘图感觉和在图板上一样习惯,绘图效率却大为提高。

#### 12.2.2 丰富实用的辅助绘图功能

IntelCAD 根据工程实际和设计人员习惯提供了丰富的辅助绘图功能。

1) 导航画对称线、平行线、垂直线、圆弧等。

2) 面向对象的智能标注和超级标注。

3) 自动捕捉加工表面的智能粗糙度标注和形位公差标注、焊接符号的标注。

4) 与使用过程密切相关的电子手册,让用户在设计中不再需要翻查手册就能获得需要的数据。

5) 与手册一样直观方便的公差查询。

6) 提供各种标准体素,如各类孔、螺纹、波浪线、辅助线、方向符号、剖视符号等,用户只需输入少许参数,系统就能自动准确绘制出所需图形。

7) 提供各种特征操作,如孔倒角、轴倒角等。

8) 相贯线绘制。

9) 倒角标注,斜度和锥度标注,以及各种形式的圆弧标注。

#### 12.2.3 可由用户灵活定制的明细表处理功能

装配图明细表及表格处理(BOM)为用户提供件号标注、明细表生成、编辑、插入、删除、文件输入输出、明细表的排序汇总,明细表的重排,重新生成和修复,旧图转换,数据库接口等功能。此外,还可以将用户已有的明细表自动处理成 InteCAD 的格式,使用户的明细表也具有 InteCAD 的这些功能。

1) 提供件号标注、明细表生成、编辑、插入、删除等明细表的标准功能。

2) 具有明细表汇总、重排及在编辑功能。

3) 可以识别读取用户明细表,充分继承用户图样信息。

4) 采用 ODBC 技术,支持对各种数据库的存取,明细表数据可以转换到各种类型的数据库中。

5) 可以由用户定制自己的明细表,符合工厂的需求。

6) 与 IntePDM 的无缝集成。

#### 12.2.4 非裁剪式装配消隐

用裁剪消隐算法实现了二维实体间的仿三维的消隐,可以任意定义、恢复实体间的遮挡关系。

### 12.2.5 功能强大的参数化设计

在 InteCAD 系统中参数化约束设计 PDAPro 模块是吸收国内外基于约束的参数化变量设计领域的最新研究成果,总结了基于约束的 CAD 系统应具备的功能要求和发展趋势的基础上形成的。在 PDAPro 命令的支持下,用户可以方便地增加、删除修改约束,以约束驱动图形。PDAPro 内核采用国际通行的变量几何求解技术,支持复杂约束模式及欠约束求解。

PDAPro 以开放性为原则,采用后参数化技术,自动识别各类约束(水平、铅垂、相切、对称、尺寸),捕捉设计意图。不论新老图纸均可直接实现参数化,为用户提供了强大的设计再利用功能。

- 1) 实时的尺寸驱动,带约束的实体或尺寸拖动
- 2) 装配方式的全参数化图形拼装。
- 3) 支持尺寸变量的代数方程组求解。
- 4) 支持机构运动分析或结构敏感度分析。

### 12.2.6 基于开放式数据库互连 (ODBC) 技术的参数化图库开发及管理工具 (PDS)

PDS 是 InteCAD 新一代的参数化图库开发及管理工具,它以新的参数化 PDAPro、开放式数据库互连 (ODBC) 技术及程序解释器 InteBasic 为技术依托,在零件参数化设计和数据库共享方面真正解决了用户的工程需要,降低了用户标准件库的开发成本。

1) 用户不用编程就建立起自己的参数化图库,库的定义过程简单统一,无论是简单库或是多级库,都只要两三个命令就可以创建。用户建立的图库运行效果和 InteCAD 系统图库的运行效果完全一致。图库可以做任意修改扩充。

2) 用户也可以将 PDAPro 和 InteBasic 在图库中联合使用,实现对复杂工程设计要求的零件的参数化定义。

3) 基于 ODBC 技术的数据库技术为用户的数据共享提供了方便可靠的保证。使用 PDS 创建的零件库,用户可以将数据存储在 Access、FoxPro、dBase 等本地型数据库中,也可以将数据存储在 SQL Server、Oracle 等网络型数据库中。在 PDS 中,对任何类型数据库(只要提供 ODBC 协议)的存取访问都是一致的。

4) 提供了一整套数据库的维护工具,用户可以浏览图库的树状结构,可以对数据库的表格进行增、删、改的操作。

5) 图库运行界面统一美观、交互性良好,允许用户随时对图库进行修改。多级库可以从任意一级开始运行,参数化零件的生成只需一两秒钟左右。

### 12.2.7 运行环境

1) 软件环境 1. Windows95 或 Windows NT;  
2. AutoCAD R13 或 R14。

2) 硬件环境 1. 微机 CPU: 80486 或 80586; 内存: 不小于 16MB; 外存: 不小于 200MB, InteCAD 软件需要的硬盘空间大约为 100MB。

3) 显示器适配卡: SVGA 卡; 分辨率: 推荐使用 800×600。

4) 鼠标。

## 13 PICAD 2000 软件产品

北京凯思博宏计算机应用工程公司 PICAD 2000 是全 32 位基于 Windows 的二维 CAD 平台,遵循 PICAD 一贯坚持的开放性、易用性。PICAD2000 面向工程设计领域 21 世纪的需求,支持多种工程设计规范 (GB 标准、ANSI 标准、ISO 标准),形成多语言的不同版本。首先推出“所绘即所得”自动参数化技术,开创了二维 CAD 技术领域又一座划时代的里程碑。PICAD 2000 作为通用二维图形处理平台是办公自动化 (OA) 系统首选图形处理软件。PICAD 带有大量国标 (GB 标准) 标准件库和标准符号库,提供常用机械零件设计计算专家系统,为加快工程设计速度提供保障。自主知识产权,完全从底层开发,无需其他系统支持,资源消耗小,对硬件要求低,运行速度快,提高投资效益。

### 13.1 全 32 位基于 Windows 的风格

1) Windows 98 风格的用户界面,风格统一,减轻用户的学习压力。

2) 多文档设计环境,在同一 PICAD 系统中同时打开任意多个图形文档,每个文档为一个窗口,在各文档窗口之间可交换各类数据,包括系统设计环境数据。

3) OLE 技术,可实现与其他 Windows 应用程序集成要求。

4) HTML 格式的在线帮助系统,图文并茂,是极具魅力的学习工具。

5) Windows 复合文档格式,支持文档部分打开,或在文档未打开的情况下,操作图形中的系统表、组合实体等信息。

6) Internet 访问支持,使 PICAD 具备通过 Internet 访问世界各地网站上文件的能力,并进一步实现全球协同工作。

7) 国际化技术,支持 Unicode,使用户利用 PICAD API 快速开发出世界各种民族语言的国际化版本,PICAD 2000 将同时推出中文版和英文版。

## 13.2 系统功能

### (1) 交互导引技术

命令行及信息窗口、导引工具条、鼠标右键上下文敏感菜单,是 PICAD 同时提供的三种交互工具,称为交互导引技术,导引用户绘图操作过程,使用户提高效率,缩短学习时间,即使是从没有接触过计算机的用户也能得心应手。交互导引技术可实现全鼠标或全键盘操作。

### (2) 历史记录管理器

PICAD 的历史记录管理器记录了用户的每一步操作过程,用户通过历史记录管理器窗口浏览历史树,预览或返回至任意历史状态,使用户真正感觉到每一次操作(包括 Undo/Redo)都是可回溯的、安全的。很多应用系统提供 Undo/Redo 功能,为用户反悔操作提供了方便,但其中的大部分系统在多次 Undo/Redo 操作工程中,会舍弃一部分历史信息,从而不能反悔至操作过程的任意状态。

### (3) 属性管理器

PICAD 2000 是真正的面向对象和面向特征设计的 CAD 系统。属性管理器是 PICAD 对象属性或特征显示或编辑的工具。它是一个可停靠窗口,窗口内以列表形式显示或编辑所选对象的属性或特征信息。显示信息将根据对象选择集的变化而变化。

### (4) 智能标注系统

标注命令将自动识别标注命令选择的对象,并按相应的标注方式标注;一次完成多个圆、圆弧的半径或直径的标注;一次完成封闭尺寸链的标注;过标注或欠标注检查。

### (5) 对象管理系统

PICAD2000 图形文件可与数据库关联,数据库连接机制采用最新的 Microsoft 数据库标准 OLEDB 2.0,实现 PICAD 对象与数据库记录的链接。PICAD 图形文件还支持数据库记录的直接存取。可选的图纸文档管理系统,对包括 PICAD 图形文件在内的各类文档进行管理,并可提取多个文档中的数据库数据,形成统一的材料表。

### (6) CAD 平台系统

PICAD 是自主知识产权的平台系统,提供应用程序开发接口 PICAD API 提供二次开发支持。开发者利用 API,可以开发 PICAD 系统的应用程序模块,在模块中定义自己的命令和实体类型,并集成到原系统中,甚至可修改或替换系统的原有功能。

### (7) OpenDWG 技术

PICAD 全面采用 OpenDWG 技术,完善保证 PICAD DRF 文件、AutoCAD DWG、DXF 文件之间的

数据交换,使 AutoCAD 用户和 PICAD 用户再无后顾之忧。OpenDWG 是一个国际 CAD 开发者联盟,目的是促进 AutoCAD DWG 文件格式的国际标准化,为 CAD 软件开发者提供最新 DWG 文件格式的转换技术。

### (8) CAD 系统环境管理工具

CAD 系统环境是个性化系统参数和系统表的总和,包括界面或操作习惯参数、计量单位、涉及到国家标准、企业标准或工作组标准的作图参数、库调用参数、图层表、字体表、颜色表、线形表、线形系数、线宽表、线宽系数、打印样式、标注样式等;每个人有各自不同的喜好,每个国家、民族、地域、单位、工作组等都有各自的规则、标准或习惯,这使不同用户需要不同的系统环境;而一个工作组内部,同一操作者在不同计算机上有时又需要相同的系统环境。在以往的 CAD 系统中,系统参数数量高达几十个,甚至几百个,学习配置或实际配置这些系统参数耗费了操作者大量的时间,且很难实现一个工作组内部或同一操作者在不同计算机上在真正统一的系统环境下工作。PICAD 2000 的 CAD 系统环境管理工具是一个数据库管理系统,为个性化系统参数和系统表提供可视化的创建或修改处理,并可导入或导出系统环境文件。它还能对基于图形的或基于系统的系统参数分别管理,并提供中国国家标准和 ANSI 标准 CAD 系统环境。

### (9) 全新的“所绘即所得”自动参数化技术

PICAD 2000 中率先实现“所绘即所得”的参数化方法,用户在绘图过程中,不须任何多余的操作,自动参数化所绘图形,不出现欠约束或过约束的情况;并同时支持尺寸驱动参数化技术,是二维图形参数化技术又一里程碑。至此,用户再也不会为没有实用的参数化技术而忧虑了。

### (10) 库管理工具

库管理工具将包含参量图符或固定图符的图形文件、数据库文件组合成 PICAD 图形库,用户可在库管理窗口中维护库,包括增加、删除、修改、导入、导出等操作,使建立和维护标准件库、常用件库更加快速、直观;PICAD 图形库支持及库管理工具能极大的提高设计效率,降低用户投资。

## 14 CIMATRON

### 14.1 CIMATRON 概要

以色列 Cimatron 软件有限公司作为世界上著名的 CAD/CAM 软件供应商一直致力于为制造行业提供先进的 CAD/CAM 系统,不仅为用户提供功能强大的通用 CAD/CAM 系统,而且针对模具制造行业提供

了全面的解决方案。Cimatron 公司成立于 1982 年,其股票已经在美国的 NASDAQ 上市。在世界范围内,Cimatron 已拥有 7000 多个用户,系统销量达 12000 套。用户广泛遍布于汽车、航空、计算机、玩具、消费类商品、医疗、军事、光学仪器、远程通信、教育机构和科研院所等。今天,Cimatron 公司在世界范围内拥有 35 个分支机构和子公司。

Cimatron 公司的 Cimatron 软件产品是一个集成的 CAD/CAM 产品,在一个统一的系统环境下,使用统一的数据库,用户可以完成产品的结构设计、零件设计,输出设计图纸,可以根据零件的三维模型进行手工或自动的模具分模,再对凸、凹模进行自动的 NC 加工,输出加工的 NC 代码。

由于 Cimatron 在中国的销售量逐年上升,Cimatron 公司决定将 Cimatron 北京代表处(1998 年成立)提升为自主经营的子公司——思美创(北京)科技有限公司。Cimatron China 在负责市场、销售、技术支持以及售后服务工作的同时,仍然为各个代理商 SGV、Unigenius Semidome 在各自区域内的市场、销售、技术支持以及售后服务工作中提供相应的支持。

随着模具制造行业的发展,对 CAD/CAM 系统提出了新的要求,今天 Cimatron 不仅是一套通用的 CAD/CAM 系统,而且针对模具行业提供了新的工具集——Quick 系列产品(快速产品浏览(QuickConcept)、快速模具分模与拔模角度分析及分型面的快速定义(QuickSplit)、快速产品设计对比分析(QuickCompare)、快速充压模具设计工具(QuickDie)等)。

今天在中国 Cimatron 的用户群也在急剧的扩大,已广泛应用于各个行业,成为当今模具设计与制造公认的首选产品,很多中国的知名企业都采用的 Cimatron。这些用户包括:春兰机械制造有限公司,海尔模具有限公司,奥克玛模具中心,美的模具制造有限公司、新科机械制造有限公司,河南新乡新飞模具制造有限公司、济南轻骑集团、济南小鸭模具制造有限公司、山东时风集团、洛阳中国第一拖拉机集团、长春客车厂——国凯模具制造和设备维护有限公司、沈阳飞机制造有限公司、陕西飞机制造有限公司、陕西西安航空发动机公司、陕西西安安泰叶片制造有限公司、北京重型电机厂、北京航天部一院、大连康发模具制造有限公司、南京跃进集团工模具厂、昌河飞机制造公司、杭州客车制造厂、上海飞翼客车制造厂、武汉神龙、上海大众集团多家分厂(如上海合众汽车配件厂、上海有色金属模具厂、上海大众配件厂)等等。

## 14.2 技术特色

做为世界知名的 CAD/CAM 系统,Cimatron 在数控编程方面一直处于世界公认的领先地位,主要体现在以下几个方面:

### (1) 基于知识的加工

这是 Cimatron 系统内置的加工核心特色,该特色保证了在用户选择了加工对象与合适的加工方法及其工艺参数之后,对加工的状况进行分析,合理地调整刀路轨迹,使加工结果更合理、更安全、更高效。

### (2) 基于毛坯残留知识的加工

做为该技术的原创厂家,Cimatron 使这一技术更加完善与丰富。毛坯残留知识是指用户根据实际加工的毛坯形状定义对应的初始毛坯(即可以定义规则毛坯,如方料、圆料等,也可以定义基于零件理论模型偏置的有形状的铸造毛坯),在加工过程中,系统时时刻刻都知道上次加工后在零件的理论模型表面上剩余的毛坯形状与毛坯特点,从而结合具体的加工方法对加工轨迹做一系列的加工优化。

1) 实现无空走刀加工;特别是在加工零件是铸造出来的时候,该特点更加使得产生的刀路轨迹没有空走,而且加工的轨迹与效率最优;因为我们可以根据铸造毛坯的情况定义根据零件理论模型而定义沿零件表面等距偏置的铸造毛坯,而且 Cimatron 系统时时知道当前在零件理论模型表面的毛坯状态,在进行等高粗加工的时候,系统会根据当前零件表面的毛坯状态进行优化计算,只有在有毛坯的地方产生加工轨迹,避免出现像没有基于毛坯残留知识的系统在进行等高开粗时出现空切的情况。

2) 实现安全的相对安全快速抬刀、移刀:为提高加工的效率,Cimatron 支持相对抬刀位置的相对抬刀方法,相对高度可以定位高于抬起位置一定的高度,如 2mm、5mm,在移刀的过程中进行刀具与零件残留毛坯的比较,在发生干涉的情况下采用用户最初定义的安全高度进行移刀;

### 3) 实现真正意义上的刀具及其夹头的干涉检查:

a. 真实的干涉检查:没有真正的基于毛坯残留知识的干涉检查的 CAM 系统是用刀具与零件的理论模型进行比较的,容易导致实际加工的碰撞,而采用毛坯残留知识的加工的 CAM 系统是用刀具与零件上一次实际加工的残留毛坯形状进行比较,使得碰撞检查更完善、更安全也更合理。

b. 自动采用备选刀具逐次加工:用户可以指定备选加工刀具列表,在第一把刀在加工特深的峡谷与沟壑发生干涉时,系统只用当前的刀具加工可以加工的区域,而接着采用第二把刀加工它可以加工的区域,再

选另一把刀加工它可以加工的区域,直到加工完成为止。在程序输出上用户可以选择按刀具分别输出不同部分的程序,也可以在程序中自动换刀而以—个程序输出。

4) 实现完整意义上的刀具载荷的分析与速率调整优化:

a) 基于切削体积:基于毛坯残留知识的加工使得系统能真正根据刀具当前的实际加工量——加工体积——进行载荷分析,而不是根据刀宽进行推测,增加了刀具载荷分析优化的科学性与准确性。

b) 基于切削角度:Cimatron 不仅能根据毛坯状况进行速率调整优化,还可以根据刀具沿零件表面的运动角度进行优化。切入材料的角度越大速率越小,切出材料的角度越大速率越大。

c) 过载分层加工:Cimatron 载荷分析与优化技术还在余量过多的情况选择分层切削的处理方式,即对残留的过多毛坯自动分多层加工完成。

#### (3) 逐深加工技术加工特深零件

在实际加工中我们会遇到许多深度较深的零件,为达到最佳的加工效果,对这些零件的加工 Cimatron 允许用户随着深度的增加渐次加大刀具的长度,对不同长度的刀具可以限制其加工的深度范围,从而避免了采用单一长度的刀具完成整个加工而由于刀具回弹所造成的加工偏差,致使加工结果不理想。

#### (4) 直观的加工结果校验

支持最后加工结果毛坯与理论模型比较,使编程人员迅速知道当前所编程序在零件实际加工后的基本状况,从而帮助编程人员进行编程决策,即知道需要进行进一步加工的情况,进行进一步的编程。

#### (5) 基于工艺特征的自动化编程

作为最早应用该技术的 Cimatron 公司,使得该技术的应用更加灵活。该技术可以把工厂的典型工艺过程与参数存储起来形成加工模式,当有新的零件可以采用同样的加工方法进行加工时,技术人员不用从头编制加工工艺,让系统根据技术人员选定的加工模式实现自动编程。该技术有以下几个特色:

1) 工艺模式不用特别定制,只要认为某个加工过程具有复用性,用户即可存入加工模式。

2) 加工模式的工艺参数表具有工艺参数的参数相关性,几个工艺参数的修改可以使其他相关的工艺参数随之修改,如用户该用另外一把刀具时,切深和侧向吃刀量等都会随之修改,这不仅增加了工艺参数定义的快捷与方便性,也增加了工艺模式的适应性。

3) 以多种方法实现几何信息的识别与提取。一个零件的加工包含多个加工步骤,不同的加工步骤中可能有些步骤只涉及零件的相关的部分,这时就要求工

艺模式能自动识别与某一加工步骤相关的对应几何信息。Cimatron 可以允许用户指定多种方法来完成,如基于颜色识别的机制、基于线型的机制、基于几何集合定义的机制以及该三种机制的结合使用。

4) 由于这些工艺结合了使用单位产品特点、编程特点以及实际设备等加工状况的大量实际情况,使得这些工艺更合理、更精练、更安全、更高效;这些典型工艺可以形成编程的知识库,新来的技术人员还可以很快利用已有的工艺库进行实际的编程工作。

#### (6) 基于斜率分析技术的一体化加工策略

1) 斜率分析技术:斜率分析技术是随着日益紧迫的加工生产需求,对加工程序编制要求更加自动化、更加迅速而产生的一个针对铣削加工特点的技术,该技术是指在程序的生成过程中,能对零件的实际形状特点进行分析与区别,从而完成特定形状采用特定加工方法的程序计算技术。

2) 一体化分析加工:Cimatron 是最早采用斜率分析技术的 CAM 系统之一,今天 Cimatron 已经实现了从单独特征区域自动分析查找技术进而采用合适的加工方法进行该特定区域加工的初级阶段,发展到了面向整个加工对象的一体化分析与加工——使用一个功能即可进行不同加工区域的识别与特点分析,然后综合整个零件的不同区域,采用最优化的、最快捷的方式对不同的区域采用不同的加工策略,同时保证不同加工区域间最佳的刀路连接与优化,保证最佳的表面质量。

3) 应用示例:如上图的清根加工,系统会自动分析出哪些区域是水平区域,哪些区域是垂直区域,在水平区域可以采用沿轮廓等距加工、平行切削加工或环行加工方式,对垂直区域采用等高线加工方式实现垂直区域的拐角清理;如左侧两个线框显示的图片所示,系统在对零件进行精加工时采用斜率分析的方法可以对整个零件或由轮廓定义的加工区域进行分析,对不同的加工区域采用更加合适的加工方法——对垂直区域采用等高线精加工,对平缓区域采用沿表面环切的方法进行精加工,从而在一个 WCUT 加工功能内即实现了整个零件或部分区域的整体精加工,同时保证了最优的加工效率、获得最好的表面质量。

#### (7) 最佳事前优化,减少事后完善—Cimatron 针对模具行业的独特加工策略

1) 最佳事前优化技术是指在模具的编程过程中,系统提供了丰富、有针对性的技术,使每一步的加工都为下一步的加工做好最佳的准备,减少由于上部加工结果不理想造成进行进一步加工编程时增加编程人员对加工条件的考虑,同时也影响加工的效果。

2) 事后完善是指在当前加工条件不良的情况下,



用户使用一些辅助的编程设定或优化方法,避免当前不良状况的影响,如避免刀具过载等。

3) 在 Cimatron 系统中的许多编程技术都对这样的要求给予了充分的考虑。比如享誉全球的 Cimatron 的 WCUT 是一个集粗加工、半精加工和精加工于一体的加工方法,使用该方法进行粗加工时,系统会在两层之间发现在平缓区域的过多余量进而对这些对精加工有不良影响的部分进行自动的层间再加工(采用逐次等高法,沿面光刀法等),从而使精加工之前在零件表面的余量尽量均匀,避免局部刀具过载发生崩刀,也使得将来的精加工时不会由于不得以的低速切削造成表面质量不够好,为下一步的精加工做好充分的准备。如果在精加工时为避免由于余量不均对刀具产生不良影响,如刀具过载,采用事后的刀具载荷分析与优化,虽然可以避免事故的发生或影响零件的表面质量,但不是最佳的处理办法。因为如果在零件的表面到处都是大余量的地方,到处都要进行低速切削,不仅影响了加工效率,而且加工的表面质量也会受到影响,特别在高速铣削加工的应用中更加忌讳这种情况的出现,高速加工也不成为高速铣削加工了。

(8) 功能丰富、完善、安全和高效的高速铣削加工在任何形式下都可以用 NURBS 插补输出来支持高速铣削;先进的高速铣削能力:螺旋进刀,圆角走刀,层间、行间的圆弧连刀和摆线式加工使走刀轨迹更加光滑以适应高速铣削;在残留毛坯知识的基础上优化进给率;通过自动调节进给速率和自动分层来确保恒定的刀具载荷,对高速铣削实现行间的圆弧连刀和螺旋式移刀等等。Cimatron 高速铣削已经在国内得到了广泛的应用,如海尔集团模具制造有限公司、春兰机械制造有限公司、西安航空发动机有限公司—安泰叶片等等。

(9) 支持紧急任务的并行编程加工模式

模具制造行业有时面临着紧迫的任务,这时如何把一个任务分布到不同的编程人员实现协同加工至关重要。由于 Cimatron 有基于毛坯残留知识加工,而且用户可以把某一加工编程结果毛坯存储起来分给其他编程人员,这样其他编程人员就可以在此基础上针对零件的不同区域进行相应的精加工编程和精细加工编程;结合 PDM 会使该项工作更便捷。

(10) 具有时时决策支持的数控编程系统

Cimatron 为用户数控编程中遇到的可能问题提供了辅助分析工具,如加工前零件可加工区域的检查(检查是否又倒角而不能加工的区域),零件最小曲率半径的大小与所在区域,当前残留毛坯的即时察看(用户不用仿真模拟就可以指导当前零件的加工结果如何,而且用户可以进行动态剖切分析),余量分析校验

等等。

(11) 以项目为中心进行技术信息管理的 PDM 系统为平台的信息共享与协作环境 Cimatron 的 PDM 系统,以客户/服务器模式为技术人员提供了一个信息集成的应用平台,使得任务的计划与进度管理成为可能,实现了以项目为中心的技术信息的集中管理,为技术信息安全提供了有效的保证。在信息管理中,不仅可以进行项目管理、文档管理、任务管理、工作流程管理和工作进度管理,所有与技术相关的其他文档都可以得到管理,如 Word、Excel、PowerPoint 等,模具模型、加工代码、技术模版和机床设备状况等也可以实现一体化管理。

## 15 ADCEUS

### 15.1 CADCEUS 概要

CADCEUS/MOLD DESIGN 软件是根据模具生产实际的经验,把工业诀窍和高性能的工程应用软件结合起来的产物。它将模垫板及零件浇注槽和孔等生成的工作变得非常方便,把模具设计师从繁琐的工作中解放出来。

CADCEUS/MOLD DESIGN 软件绝对是针对模具设计师特殊所需面开发的专用工具。从零件设计,尺寸标注,模腔模具设计,直到制造和成品控制等各阶段,CADCEUS/MOLD DESIGN 都能完全按照具体要求,为用户提供全面的塑料工程解决方案,满足模具设计工业的各种需求。

### 15.2 零件设计

- 1) CADCEUS 主体架构是百分百的综合模型系统,免除曲面跟实体混合时所出现常见的弊端。
- 2) 按着参数定义可随着转换或更换。
- 3) 提供直接及标准的接口格式,以便跟不同的软件沟通。
- 4) 提供强劲薄壳肉厚的功能(可接受不同壁厚)。
- 5) 自动拔模倾斜角计算。
- 6) 全自动整体倒角生成。
- 7) 具备收缩补偿(等值,不同 XYZ 缩水等)。
- 8) 多种功能作模芯/模穴,分模线/分模曲面的形成(无论是平坦或曲面均可)。
- 9) 自动凹割(under-cut)检测。
- 10) 还原及重做功能。
- 11) 配备完整模具设计流程的功能。

### 15.3 模具设计

- 1) 三维参数化模具零件库。



- 2) 只需输入足够参数,可快速形成体体模具的配置。
- 3) 能在三维环境下,快速地将所选的零件定位及进行装配。
- 4) 全功能二维绘图,包括完整尺寸标注。
- 5) 二维及三维模组互交替地应用及同步更变,保持相交性。
- 6) 动态组装及预检测功能。
- 7) 零件采购单(BOM)可以 CSV 或 EXCEL 格式输出
- 8) 具有冷却系统及流槽浇口。
- 9) 在所需的公差,标准零件如滑动模芯,嵌件,顶针等自动生成。
- 10) 用户可自行设计非标准零件及模具垫板。
- 11) 自动干涉检查(冷却系统 VS 螺栓)。

- 12) 零件尺寸输入按库存显示。
- 13) 支持 Futaba, Punch, Misumi 等规格。

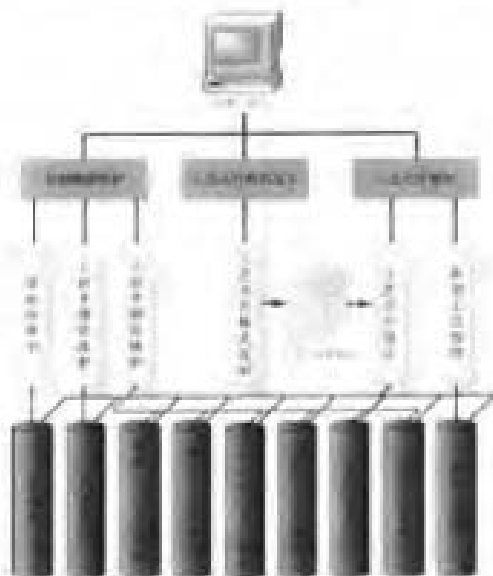
#### 15.4 工模制造及生产

- 1) 加工总汇管理系统。
- 2) 支持高效率,高生产力的 2 轴,2.5 轴及 3 轴加工机能。
- 3) 可作模芯/模穴/分模曲面/钢板材/EDM 电极的加工。
- 4) 钻/镗/铣削零件及冷却系统定位孔洞。
- 5) 自动产生曲面交线清角加工路径。
- 6) 能免除撞刀情况出现。
- 7) 能按着毛料变更的管理,自动确认残余物料及其加工路径。
- 8) 复合加工。

## 第 2 章 CAPP 软件

### 1 SIPM/CAPP

思普工艺设计与管理系统 SIPM/CAPP 由思普公司开发,是国内主流商品化 CAPP 软件中最著名的平台产品之一。它不仅在提高工艺设计效率,减少工艺设计中的重复劳动方面具有优异的功能,而且在实现工艺设计文档电子化和管理的科学化,实现企业工艺知识的积累,提高工艺数据的利用率,提高工艺标准化程度,降低成本,保证质量,以及直接应用设计资源,与 PDM 系统双向信息集成等方面都有非常好的表现。是一款从企业工艺信息化整体着眼的优势 CAPP 软件产品。SIPM/CAPP 的系统结构如图附-2-1 所示。



图附-2-1

#### 1.1 软件功能

SIPM/CAPP 系统在国内率先以工艺设计工具平台的方式推出,彻底解决了 CAPP 适应性差、实用性差的积弊,成为适用于所有行业、所有工种工艺设计与管理和格式化数据处理需要的商品化的 CAPP 平台软件,能够为用户提供所见即所得的工艺卡片格式定制,工艺卡片设计,工艺文件电子签名,工艺文件提交归档,工艺文件 Web 发布与浏览,工艺统计与报表管理,工艺设计资源库管理,工艺设计知识库管理、典型工艺库管理,成组工艺设计,工艺构件开发与集成等完整工艺设计与管理的平台化解决方案。

##### (1) 所见即所得的工艺设计界面定制工具

SIPM/CAPP 提供所见即所得的工艺设计界面定制工具,企业可以根据自身需要开发个性化的 CAPP 系统,设计出 CAPP 的整套工艺设计工作界面,通过这些界面可以一次性地完成工艺信息的采集、组织、存储和报表过程,使工艺设计过程不再人为地割裂为数据采集、报表等过程,工艺设计能力也不再依赖于软件提供的初始功能,而是可以无限制地调整和扩展,从而能够满足企业快速变化的业务需要。

工艺设计界面定制工具还将工艺标准化工作与工艺设计工作紧密地融为一体,由于采用了所见即所得的技术,工艺设计工作界面即是标准工艺卡片格式,这使得工艺标准化在 SIPM/CAPP 软件中达到了形式和内容的完全统一。

##### (2) 简易高效的工艺设计环境

基于 SIPM/CAPP 形式和内容完全统一所见即所得的工艺设计界面,工艺设计师可以利用系统提供的一系列工具轻而易举地进行工艺数据的关联填写和关联修改,可以在工艺知识资源库的支持下自动智能地提示可供选择的工艺数据,可以方便地复制重用相似的零件级工艺和工序级工艺,可以简便地获取 PDM 中的设计数据和 CAD 图形数据,可以通过过程卡派生自动生成对应的工序卡片并继承过程卡中已有的信息,可以实现文字和图形的完全混排,可以根据需要自动压缩显示表格单元中过长的工艺内容,可以直接地定义编辑和修改形位公差,粗糙度,尺寸公差,分数等形式的工艺特殊内容,可以自动实现工艺格式特有的文字编辑和排版,可以动态地管理工艺文件中所包含的所有工艺卡片页并输出相应的工艺卡片目录清单,可以在工艺卡片页管理树的导航下进行工艺设计内容的快速定位和切换等。总之,工艺设计中的重复劳动得到最大限度的消除,同时,人性化的使用界面和经过优化的操作步骤更贴近工艺设计人员的传统习惯,使工艺设计的效率倍增,设计环境如图附-2-2 所示。

##### (3) 工艺文件电子签名模块

电子签名是 SIPM/CAPP 系统的特有功能,它是以真实笔迹或章戳的方式来显示和控制工艺文件的生命周期演化过程,这些业务过程包括工艺的设计、校对、审核、会审、批准、局部修改、发布、废止等。

##### (4) 工艺管理模块

SIPM/CAPP 自带工艺管理模块,工艺设计人员在设计环境中可以完成从设计、提交、签审、归档、查



图例-6-3

询、变更（投入/撤出）等完整的业务过程，使相关的工艺工作可以在同一界面下一气呵成。

同时，灵活的系统架构又使得 SIPM/CAPP 能够平滑自然地接受各种 PDM 系统的管理，而没有管理上的鸿沟，其支持的业务过程与采用 SIPM/CAPP 的自带管理模块无异。

#### (5) 工艺文件 Web 发布与浏览工具

工艺文件 Web 发布与浏览是协同工作环境下的必备工具，SIPM/CAPP 提供两种工艺文件的 Web 发布与浏览工具或方式，一种是以 PDF 格式发布与浏览，第二种是以 Web 插件的方式支持网络发布与浏览。

#### (6) 工艺统计与报表管理工具

SIPM/CAPP 中内置有工艺管理模块和产品结构 BOM 引擎，其中产品结构 BOM 引擎包括本地 BOM 和远程 BOM 两种方式，分别与自带管理模块和外部管理系统（如集成的 PDM 系统）相对应，工艺统计和汇总等功能均基于产品结构 BOM 进行，统计和汇总的内容包括工装、设备、工艺路线、工时、自制件、外购件、外协件、材料消耗定额等多种视图，统计和汇总

结果可以按照用户自定义的任何复杂格式进行报表输出。

事实上 SIPM/CAPP 系统强大的表格数据处理引擎已经远远超出其作为 CAPP 软件的传统作用，它甚至可以被集成到任何数据库管理系统中作为报表工具使用，典型应用实例包括惠普产品数据管理系统 SIPM/PDM 的报表系统。

#### (7) 工艺知识资源库管理模块

SIPM/CAPP 的工艺知识资源库包括基础资源库、工艺知识库和工时计算库等组成部分。

**基础资源库：**对工艺设计中用到的基础数据进行描述，并将这些基础数据如，部门信息、冷却液、工装、设备、工艺用语等信息整理入库，使得工艺设计人员在工作中能够方便查找调用。基础库不仅能够提高工艺设计效率，同时还能提高企业工艺设计规范化、标准化程度。

**工艺知识库：**系统配备的工艺知识库涵盖各类国家标准工艺手册、企业标准及行业标准，系统可保证工艺设计人员能够在庞大的工艺知识库方便快捷地检索到所需的工艺信息。

工时计算库,系统提供用户自定义的工时标准管理框架,使企业轻松建立适合自身的工时计算库,为工艺设计提供快速准确的工时定额计算。

系统支持工艺知识资源库的开放性和完全可扩充性,设计人员可以在系统提供库的基础上根据自己的

需要,建立个性化的工艺知识资源库,加入个性化的内容和规则,从而使得个人使用的数据库更加简洁,更有针对性,使用效率更高。

基础库和工艺知识库分别如图附-2-3、4所示。



图附-2-3 基础资源库

#### (8) 典型工艺库管理与成组工艺设计模块

典型工艺管理模块(图附-2-5)是SIPM/CAPP的一个重要高级模块,它为企业提供了一个根据零部件工艺相似性管理和重复利用工艺数据的工具平台,有助于企业不断积累工艺知识和工艺数据,并充分利用已有工艺设计成果,进行快速工艺设计,从而显著地提高工艺设计效率,工艺规范化和工艺设计质量,避免出现重复的质量缺陷。

在典型工艺库的支撑下,能够根据零部件的工艺相似性参照已有的成熟工艺,辅以高效的编辑修改操作,快速完成新工艺的设计过程。

#### (9) 工艺构件开发与集成工具

SIPM/CAPP灵活的系统架构又使它能够平滑自然地接受各种PDM系统的管理,而没有管理上的鸿沟,从与SIPM/PDM、JMAN等商业PDM软件的已有集成案例来看,其集成可达到双向无缝的程度。

SIPM/CAPP系统提供集成中间构件库,使任何PDM都只需简单开发即可轻易实现对SIPM/CAPP工艺文件和工艺信息的管理,在工艺设计环境中对工艺文件和工艺信息进行在线的查询、统计、汇总和归档,亦可在PDM环境中浏览、查询、统计、汇总乃至进入工艺设计环境进行工艺设计和工艺变更。

## 1.2 SIPM/CAPP软件的特点

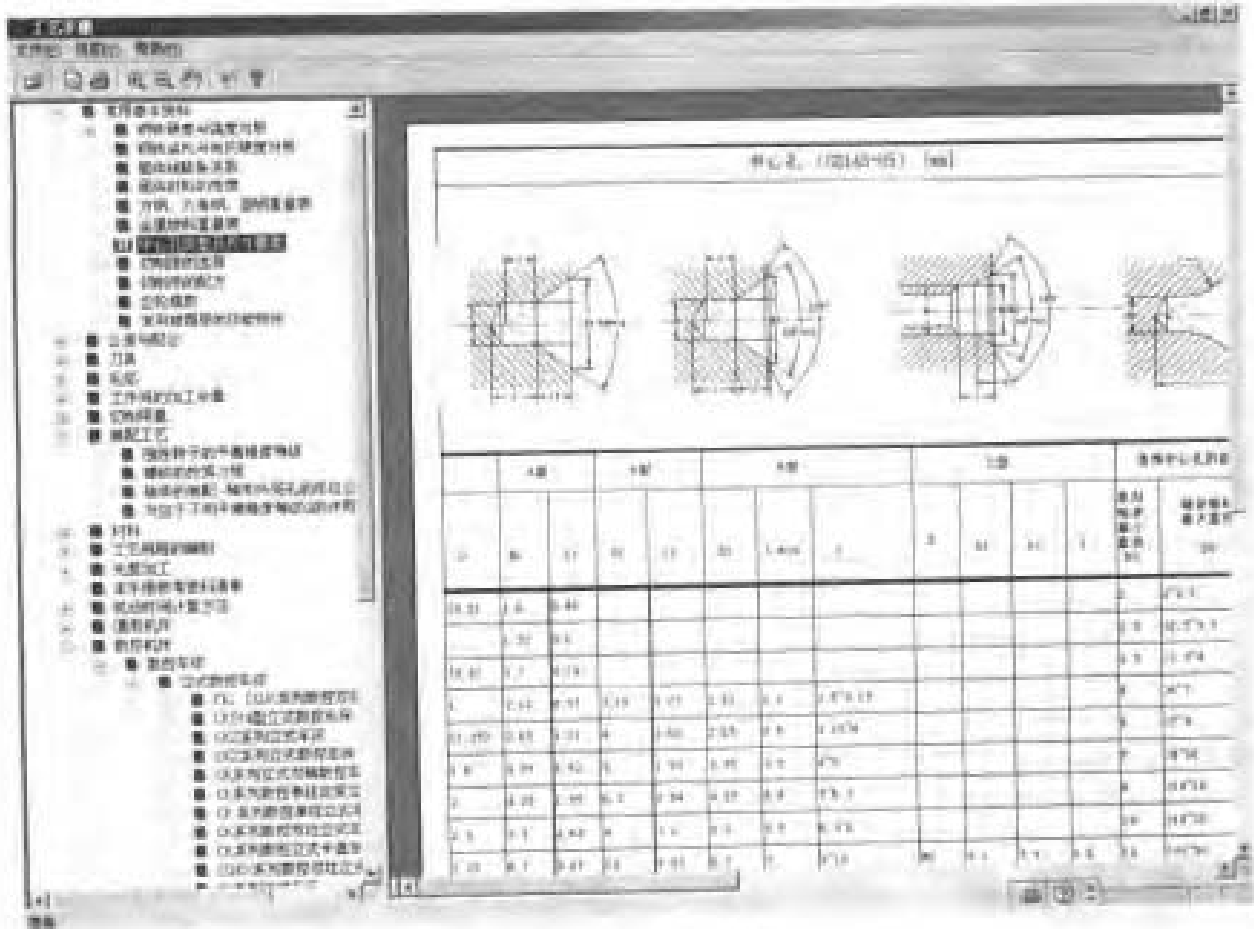
该软件的特点有:

(1) 界面友好、操作简便、实用性强

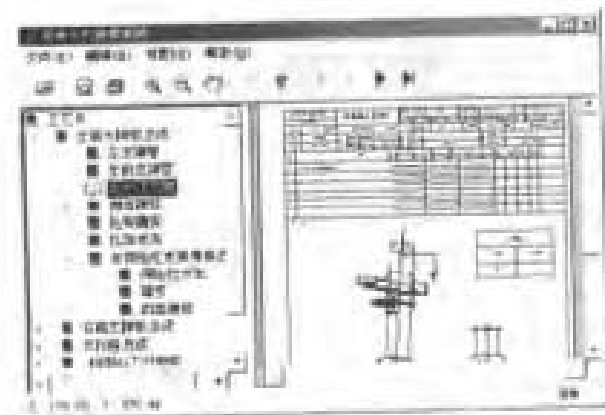
1) 友好的用户界面,易学易用,两天即可完全掌握使用技巧。

2) 完全所见即所得,设计、组织、存储、排版、报表等一次完成。

3) 多文档支持,在多个设计任务之间可以切换自如。



图附-2-1 工艺知识库



图附-2-2 典型工艺管理系统

4) 基础数据在线智能提示, 令用户的工作事半功倍。

5) 卡片树管理树导航, 工艺文件结构一览无余, 调整方便。

6) 典型工艺管理支持, 成批设计效率倍增。

7) 关联填写和关联修改等功能最大限度地取消工艺设计中的复杂劳动。

(2) 功能体系完整

SIPM/CAPP 系统功能覆盖了标准工艺格式制

定, 工艺设计, 签审流程, 工艺管理, 工艺统计汇总与报表, 工艺知识资源管理, 典型工艺管理, 开发工具与集成, Web 发布与浏览等完整的工艺业务流程, 用户可以通过它完成几乎所有的相关工艺工作。

(3) 强大的可扩展能力

SIPM/CAPP 提供了强大的可扩展能力, 主要包括:

1) 设计和定制个性化工艺设计界面, 利用工艺界面定制工具可以非常简便地任意绘制企业的标准工艺卡片格式和个性化的工艺设计界面, 从而构建完全属于企业自身的 CAPP 系统。

2) 扩展内置的工艺模型: 用户可以根据企业实际情况自行扩充工艺模型的内容, 以适用企业更加复杂的工艺描述要求, 系统在设计过程中能够自动生成扩展后的工艺模型, 并支持模型的自定义数据输出。

3) 扩展工艺知识资源库, 包括基础数据库、工艺知识库, 工时计算库等在内的工艺知识库均可以无限扩充和扩展结构模型, 系统将自动识别并利用扩展的工艺知识资源库。

4) 扩展自定义报表: 通过系统提供的报表配置工具, SIPM/CAPP 能够将数据输出到企业自定义的格

式化表格中,实现报表功能的动态扩展。

#### (4) 优良的可集成性

1) 能够直接利用包括 AutoCAD、SolidEdge、SolidWorks、UG、Pro/ENGINEER 等 CAD 软件绘制的图形。更可以在思普工艺系统中直接打开 dwg/dxf 文件作为工序图,颜色、粗细、大小可以进行任意调整,保证工序图符合标准要求。

2) 通过简单配置可以直接实现 CAPP 与各种 PDM 的数据集成。

3) 通过提供的工艺集成构件进行二次开发,可以无缝地接受各种 PDM 的管理并实现功能集成。

4) 作为通用的数据报表工具与数据库管理系统集成。

## 2 Inte/CAPP

武汉华中软件公司 Inte 系列的 InteCAPP 系统产生的工艺卡片可以纳入 IntePDM 系统的管理,与 InteCAD 无缝集成。采用网络数据库 (Windows NT/SQL Server 等);便于填写表格、利用 CAD 图形数据、保证数据正确性、信息共享,实现并行设计。

Inte/CAPP 基本功能有:

#### (1) 工艺项目管理

1) 建立工艺项目 (基本件、通用件、标准件、配套件、电器件、图样等)。

2) 工艺项目预处理:从 PDM 提取产品设计数据,区分“领用件”与“自制件”、确定工艺设计的各个部分分别由哪些模块完成。

#### (2) 工艺生成子系统

检索式工艺生成子系统。按照系统提供的检索策略和方法从工艺数据库中检索出与被设计零件相似的工艺过程。再利用“编辑式工艺生成子系统”对所生成的工艺规程进行编辑修改,使符合用户要求。

派生式工艺生成子系统。根据被加工零件的工艺特征 (工艺编码),查询对应零件族的样件工艺,再加以修订而成,即所谓派生。企业根据多年的工艺积累大多已总结了许多类零件的指导性工艺,甚至已有采用 GT (成组技术) 概念建立零件族样件工艺基础。该方法有较好的理论和应用基础。通过不断整理零件族的分类编码和标准工艺,成熟一类,加入一类,以逐步扩大修订式工艺生产的覆盖面。

编辑式工艺生成子系统。“编辑式工艺生成子系统”与用户交互进行工艺设计,提供用户快速制定和输入新零件工艺过程的手段。提供方便、快捷的工艺过程编辑器,丰富的工艺文件编辑手段;提供丰富的工艺知识数据库,用户在设计过程中可以迅速准确地查询使用有关工艺知识和数据;用户制定工艺过程的工作中,

系统会自动向用户推荐相关工艺知识与数据。各个工艺生成子系统最终都会进入到“编辑式工艺生成子系统”的编辑界面上进行最后的调整。在“编辑式工艺生成子系统”中还包括了“工艺校对”。

#### (3) 辅助工艺处理

辅助工艺处理是集成于 InteCAPP 平台的一系列应用程序。它们依靠 InteCAPP 平台的统一数据环境,完成其他工艺设计工作。其内容包括:编制指导工艺卡;编制装配工艺卡;辅助工艺计算。进入工艺计算环境,完成各种特定的工艺计算,如工时定额。工时定额制定的方法可将算法与经验法相结合,系统适时提供相关定额图表、同类工艺档案供参考。

#### (4) 工艺设计后处理

操作临时“工艺项目数据库”,对生成的工艺文件进行汇总,用户对工艺文件进行调整性处理;输出最终的工艺 (项目) 文件至“MIS 数据库”和“工艺项目保存数据库”中;按工艺设计项目的产品结构树所表示的装配关系,对自制件根据选择毛坯自动分类汇总所有下料清单。这个工作人工操作十分繁琐又易出错,却正好是计算机的强项。可自动生成工艺路线卡片,工艺卡片目录,工时定额卡片;集中 (网络打印或绘图) 输出制造系统所需的全部工艺文件的硬拷贝。如:工艺过程卡、下料清单、工艺线路卡片,工艺卡片目录,工时定额卡片,指导工艺卡,装配工艺卡等。用户从产品结构树中选择一个分枝,系统可以生成其工艺卡片目录,打印输出其中所有的工艺过程卡。

#### (5) 工艺数据库的维护

对工艺数据库进行适时的维护和管理。工艺数据库提供了丰富的工艺数据,包括机床数据、刀具数据、夹具数据、量具数据、材料数据、加工步骤描述术语、工艺符号等,较完整地收集了《机械加工工艺手册》的有关内容。工艺数据库向用户开放,系统支持用户对工艺数据库进行修改和扩充。

#### (6) 定义卡片格式

模板图是根据工艺卡片的格式用 AutoCAD 绘制,并对应卡片上每个需填充的区域编号的卡片图。系统根据用户定义创建卡片子表,用以存储卡片上各个子表中的工艺数据。

#### (7) 明细表处理功能

1) 明细表的数据库功能。能够将明细表内容输出到网络数据库中,自动生成指定部件的零件清单和外观件清单,可以与 IntePDM 双向交换明细表和标题栏数据。

2) 可以读取零件定义系统 PDS 生成的标准件的产品数据到明细表。

### 3 开目 CAPP

开目 CAPP 是企业级的工艺设计与工艺管理解决方案,真正实现了基于知识决策的智能化工艺设计。开目 CAPP 提供了完全基于数据库的大型工艺集成设计与管理环境,具备强大的处理复杂工艺数据表格的功能,提供了严格的工艺权限管理、工艺流程管理功能,支持团队并行工艺设计,对于主流 CAD、PDM、ERP 的产品具有良好的集成性,并提供了丰富的二次开发接口和工具,是业界公认的功能最强、商品化程度最高的 CAPP 产品。目前在国内拥有近千家用户,是市场上绝对领先的产品。

软件特点:

1) 最早提出并实现 CAPP 工具化思想,提供功能强大的、开放性好的功能组件,可以任意创建工艺表格,快速拼装 CAPP 系统,实施周期短。

2) 融数据库、图形、图像、表格、文字编辑于一体,图文并茂,提供可视化的工艺设计环境,100%所见即所得,国内领先。

3) 独创的所见即所得标注特殊工程符号技术。全面支持尺寸偏差、粗糙度、形位基准、形位公差、加工面符号等特殊工程符号,并可快速新增其他用户所需的专用符号。编辑特殊工程符号如同编辑一般文字一样,可以剪切、删除、复制、粘贴、插入等,实时显示。

4) 独有的公式计算和公式管理器功能,提供材料定额计算和工时定额计算公式库。用户可自行扩充专用公式。系统可自动筛选公式,并将计算的结果自动填入到工艺文件内。

5) 开放的工艺资源管理库。基于网络、数据库环境,支持 ACCESS、SQL\_SERVER、ORACLE 等,工艺资源的结构和数据,用户可以自己定义、扩充,可管理的数据类型包括数据表、图形、图表数据等。

6) 大量丰富、实用、符合国标的工艺资源数据库。包括设备、标准刀具、标准量具、标准工艺术语、材料牌号、材料规格等,由专门从事工艺设计和工艺管理的人员建立,简洁、实用、覆盖面广,并继续不断丰富内容。

7) 真正起到了工艺设计的桥梁作用。向上,可以直接读取各种 CAD 的设计图样信息和图形信息;向下,可以直接提供 PDM、ERP 系统所需的材料、工艺路线、工艺装备、工时等工艺设计信息,实现了真正的信息集成。

8) 独有的角色批阅和签字功能。支持 PDM 系统的设计、审核、批准等工作流程,实现了不同的角色对同样的工艺文件具有不同的权限功能。

9) 完全自主版权,开放和开发了多种二次开发接

口。可为用户快速提供专用接口,开发接口可被其他应用系统直接调用,提取出各种工艺信息。

10) 独有的打印中心模块,实现工艺文档的集中、排版输出。自动排版,既节约纸张,用实现了工艺文档输出的集中管理。

### 4 高华 CAPP

高华计算机辅助工艺设计系统采用“技术先导、实用为本”的应用策略,以交互式工艺设计为基础,实现工艺与管理计算机化,提高工艺人员的工作效率。

系统主要功能如下:

1) 工艺设计:工艺人员在标准图形界面下以数据定义方式进行工艺路线(分工路线、车间工艺路线)、材料定额制定等各类综合工艺设计和机加、装配、焊接、钣金冲压、热处理、铸造、锻造、返修等各类专业工艺设计。

2) 工艺卡片自动生成与编辑:系统根据工艺人员定义的工艺数据,在已定义好的工艺卡片格式文件支持下自动生成工艺卡片。

工艺管理:系统不仅可以对工艺数据和工艺文件进行层次化查询,而且可完成工装设备、材料、工艺关键件、外协外制作、标准件、辅助用料、关键工序等各类用户自定义的统计汇总功能,并自动生成汇总统计报表(明细表)。

3) 工艺知识库管理:通过系统提供的知识库管理功能,用户可以针对企业的实际情况建立工艺信息模型,在此基础上,用户可以用层次化方式建立自己的机床设备库、刀具库、夹具库、量具库、切削参数库、材料库、典型工艺(工序、工步、工艺术语等)库等,为工艺设计提供支持。

4) 工艺卡片格式编辑:用户可方便地以图形交互方式对企业工艺文件中的各类工艺卡片格式进行定义、修改与编辑。

5) 工艺卡片动态知识获取:工艺员在工艺设计过程中,可随时将产品工艺中所定义的工序、工步、设备、工装等数据对象不经过任何修改或经过一定的编辑修改直接放入知识库,从而实现知识库的动态扩充。

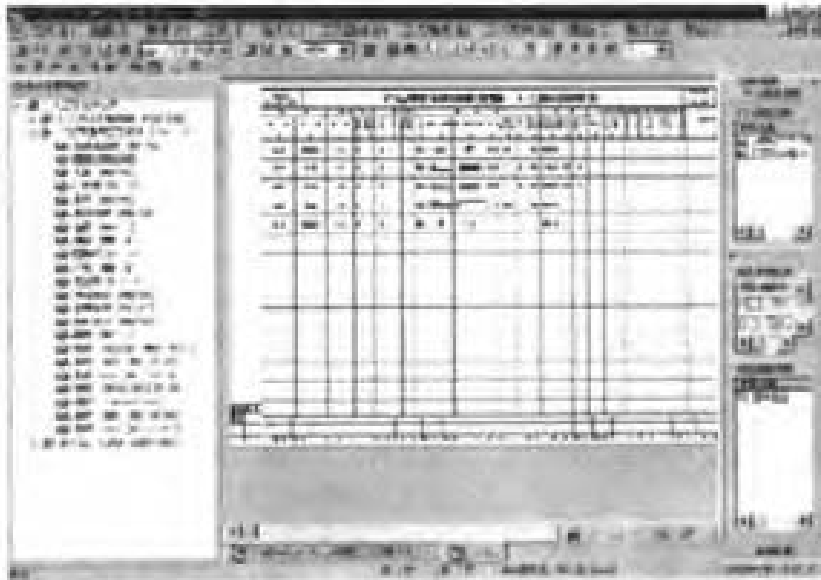
### 5 XTCAPP

XTCAPP 是北京艾克斯特科技股份有限公司最新推出的工艺数据管理系统,它全面面向工艺部门和工艺人员,兼顾了工艺编制和工艺数据管理两方面要求,既能满足车间工艺人员“甩钢笔”的要求,又能适应整个工艺部门全面管理的需要,同时也为企业信息化集成奠定了良好的基础。

功能特色:

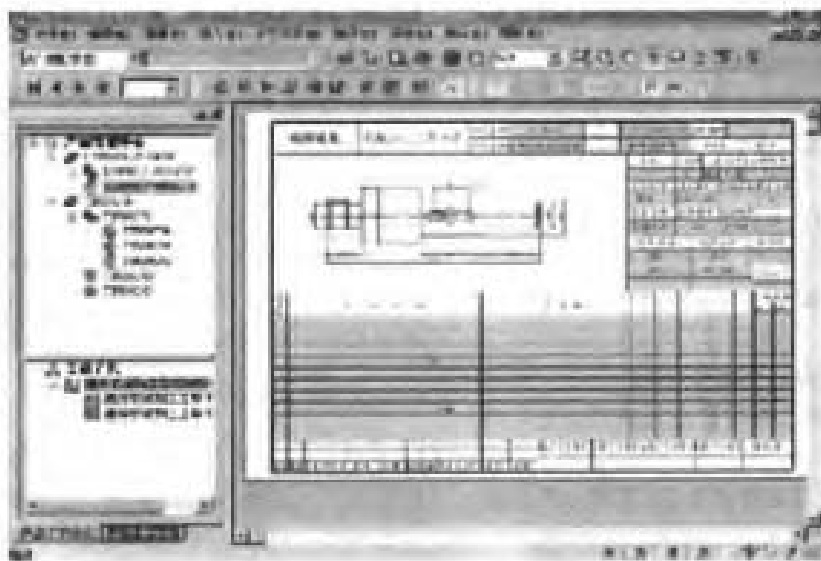
1) 基于产品结构实现有效的数据管理, XTCAPP 是国内第一个采用基于产品结构树组织和管理工艺技术文件的 CAPP 系统, 提供基于产品结构树的综合管理平台, 将工艺文件按产品结构树的方式有机的联结

起来, 同时提供方便的查询工具, 系统支持多种建立产品结构树的方式, 如 XTPDM 导入、DWG 文件读取、TXT 文件导入以及手工建立等。



2) “所见即所得”的编辑界面让工艺人员轻松工作。“所见即所得”的工艺卡片编辑模式, 与 Office 类似的操作方式, 让工艺人员在习惯的界面上完成工艺卡片的编制。系统支持各种特殊符号和工程标注, 表格

与数据库的关联体系自动对应相关的工艺基础数据最大限度的减少重复劳动, 各种便利的编辑工具和方法, 可以快速准确的生成所需的工艺文件。





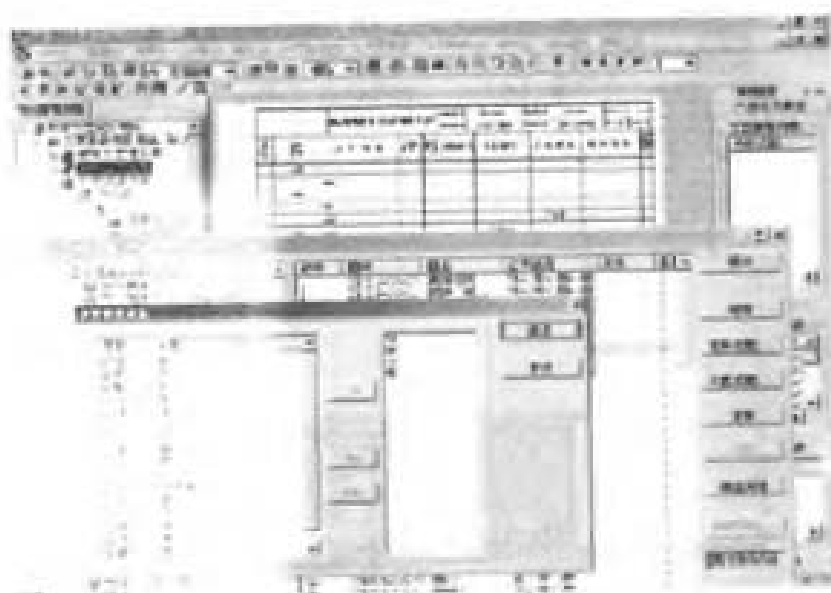
3) 开放灵活的数据统计汇总功能大量减少手工操作,支持多种格式的报表编辑及输出定义,用户自定义的报表生成数据配置功能,能够满足几乎无限的多种数据的报表生成需要,及时有效的支持企业的决策。

系统提供向导式报表定义工具和超凡的公式管理工具,为产生各式各样的报表与数据汇总提供便利的条件。系统生成的报表可再次编辑和汇总。



4) 基础数据管理整合企业资源,系统预置多套基础数据库,覆盖企业绝大多数的基础数据的维护和查询需要,同时又具备无限的扩充能力,能够迅速满足企

业对其他的基础数据的要求。工艺系统同时在此模块的基础上实现标准工艺库的功能,提高企业工作的效率和数据的规范化。



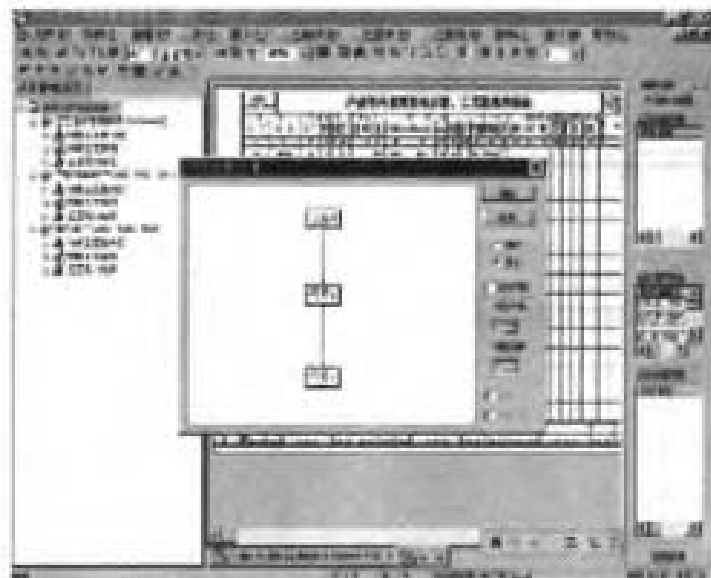
5) 开放的典型工艺管理,减少重复工作:系统提供开放的典型工艺管理功能,以分类结构方式有效管理企业宝贵的工艺文件数据,提供产品级/部件级/零件级

的工艺文件重用,对于系列产品的工艺设计,能够最大限度的减少重复输入和提高工作效率。



6) 严格周密的用户管理保证企业数据的安全:按照“组织”和“角色”来组织和管理用户,可以对工艺数据、基础数据等进行分别授权,保证各种工艺数据的

安全,同时系统提供全面的日志管理功能,详细记录用户对工艺数据的操作,实现数据变迁的可追溯性。



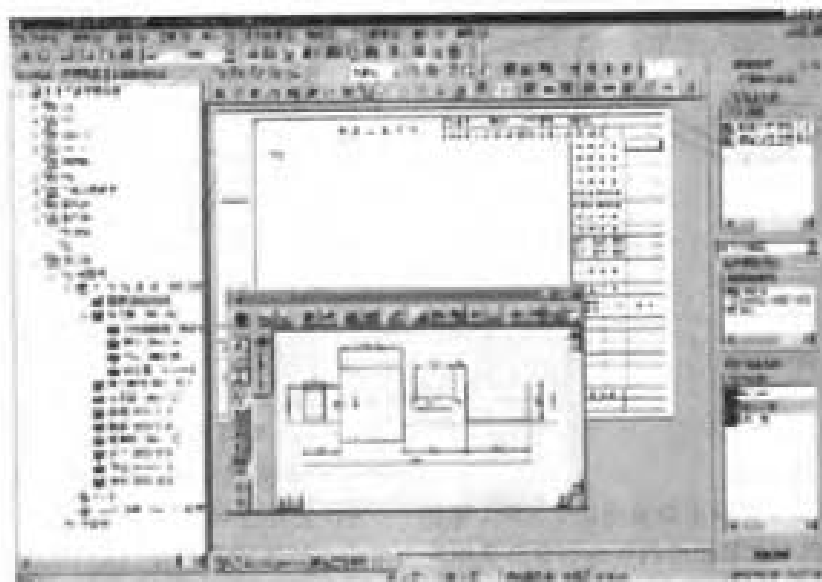
7) 灵活的图册管理大幅提高工艺卡片输出的效率,系统提供工艺图册管理,可以按照产品或者工艺文

件种类等不同方式将工艺文件进行集中排版和优化,并支持大幅面绘图仪的集中输出,提高出卡效率。



8) 与 XTPDM 紧密集成,保证数据唯一: XTCAPP 标准版可与 XTPDM 紧密集成,直接读取和

使用 XTPDM 的设计视图和产品结构进行工艺设计,保证设计数据的准确传递。



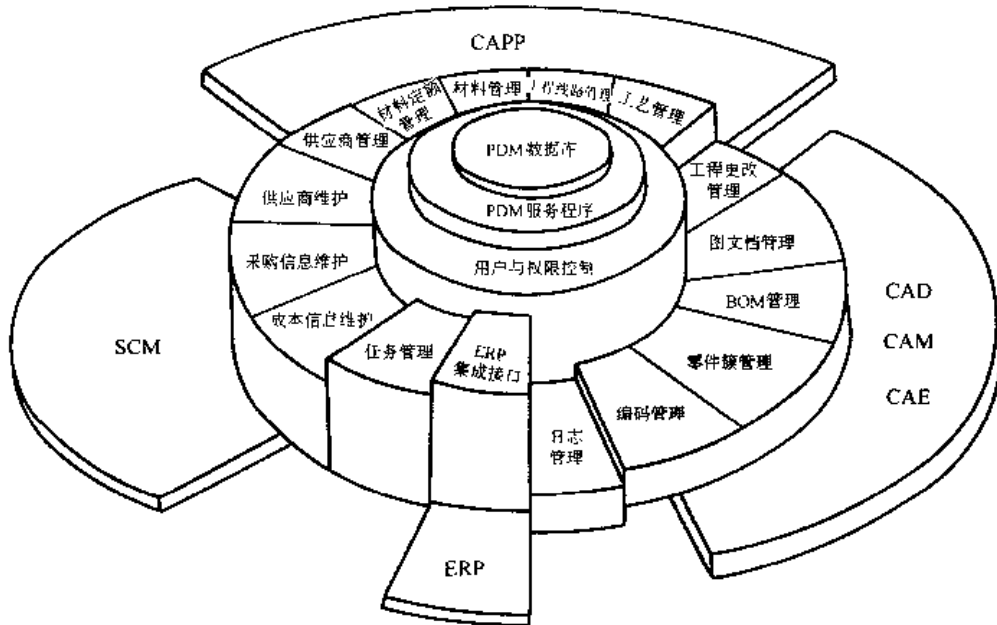
# 第3章 PDM 软件

这里介绍 IMAN、Windchill、CATIA/PDM I、Metaphase、SIPM/PDM、IntePDM、BEE-PDM、CV/Optegra、InSync 和 SherpaWORK 软件。

## 1 SIPM/PDM

SIPM/PDM 是由思普公司开发并于 1999 年首次推出的产品数据管理软件，目前已经发展到 5.0 版本，产品推出后陆续在汽车、航天、航空、发动机、工程机

械、家用电器、印刷机械、橡胶轮胎等多个制造领域得到成功应用，著名的用户包括中集集团、上海大众、上海通用、东风股份、长安铃木、汇众汽车、德尔福汽车空调、采埃孚转向机、上海航天局、威孚高科股份、临工股份、宣工股份、锡压股份、橡胶轮胎股份、南车四方股份等众多大型企业或企业集团。SIPM/PDM 的基本结构如图附-3-1 所示。



图附-3-1

### 1.1 软件功能

SIPM/PDM 覆盖了产品生命周期的全过程，管理分布于全企业范围的与产品相关的所有信息及其相关的所有活动，包括从产品设计到产品出图、从工艺分工到工艺设计、从产品装箱发运到售后服务等环节在内的整个过程及有关的信息，协助技术部门并行协同地完成复杂的产品开发项目。根据企业的不同需要，SIPM/PDM 提供了众多的可选功能模块。图附-3-2 是 SIPM/PDM 的集成主界面。

SIPM/PDM 主要模块如下：

#### (1) 零件簇管理

- 1) 按照零部件相似性管理所有零部件信息及相关资料。
- 2) 实现零部件的分类化管理，支持多种分类方式并存。
- 3) 建立、编码零部件并维护其属性。
- 4) 管理零部件的互换性和替换件。
- 5) 快速查找定位零部件信息。
- 6) 根据零部件查找图样、资料。
- 7) 对零部件进行通用度评价。

#### (2) BOM 管理



图附-3-2

- 1) 实现产品系列管理和分类管理。
  - 2) 建立和维护产品。
  - 3) 建立和维护产品版本。
  - 4) 建立和维护产品视图。
  - 5) 按照产品及其组织结构管理与产品有关的所有数据。
  - 6) 建立和维护产品结构，及其相关组成零部件。
  - 7) 派生设计 BOM 结构。
  - 8) 建立、管理和维护借用关系。
  - 9) 快速确定设计更改影响范围，并实施一致性自动变更修改，记录变更历史。
  - 10) 比较产品及零件 BOM 的结构差异。
  - 11) 结构相同性比较。
  - 12) 物料快速定位。
  - 13) 物料属性可配置定义。
  - 14) BOM 视图转换，管理设计、制造等 BOM 多视图及其关系。
  - 15) 追溯设计变更历史。
  - 16) 产品级和部件级的 BOM 卷积计算、各类统计、汇总与报表生成。
- 图附-3-3 为产品 BOM 的报表示例。

- (3) 产品配置管理
  - 1) 产品配置模板定义及模板配置。
  - 2) 变量配置规则定义及变量配置。
  - 3) 参照配置及基于设计 BOM 演化各种制造 BOM，如工艺 BOM、生产 BOM、采购 BOM、装箱 BOM 等。
  - 4) 自由配置环境设置及自由配置。
  - 5) 产品数据的一致性校验。
  - 6) 产品数据的完整性检查。
  - 7) 由过程数据到归档数据的转换。
- (4) 图档管理
  - 1) 按照产品结构和零件族等多种方式管理图档资料。
  - 2) 根据权限在线阅读各种图档资料。
  - 3) 通过浏览器比较 DWG 文件图形，如一张图样从另外一张图样修改后得到，可直接显示其差异部分的内容。
  - 4) 过程文档、正式文档、历史文档等与生效有关的过程控制。
  - 5) 版本控制。
  - 6) 检人检出机制控制团队设计协同。

- 7) 按照更改通知自动触发图样的生效状态更改。 生成产品结构并关联入库对应图档。
- 8) 批量提取 DWG 图档的标题栏和明细栏信息。 图附-3-1 是 SIPM/PTM 的图档管理界面。

上海通用汽车 设计开发系统		C) 物料结构				物料信息		物料属性		物料位置				物料状态	
物料号	物料名称	物料描述	物料规格	物料单位	物料数量	物料属性	物料属性	物料属性	物料属性	物料属性	物料属性	物料属性	物料属性	物料属性	物料属性
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

图附-3-1



图附-3-4

**(5) 工艺管理**

- 1) 按照产品结构管理工艺文件。
- 2) 支持 SIPM/CAPP 工艺文件在线浏览。
- 3) 工艺文件版本管理与生效控制。
- 4) 过程文档、正式文档、历史文档等与生效有关的过程控制。

- 5) 检入检出机制控制团队设计协同。
- 6) 支持按照产品结构统计和汇总工艺信息。
- 7) 材料消耗定额的维护。
- 8) 材料消耗定额的更改控制。
- 9) 材料消耗定额明细表、汇总表及格式化报表输出。

10) 材料消耗定额更改记录。

- 11) 工艺分工路线的编制。
- 12) 工艺分工路线的更改控制。
- 13) 按部门提取工艺路线。
- 14) 工艺路线全局替换。

**(6) 编码管理**

- 1) 建立企业级的物料编码规则。
- 2) 在 BOM 管理中进行交互编码。
- 3) 在零件簇管理器中进行交互编码。
- 4) 可用于所有需要规则编码的场合。
- 5) 回收废弃编码。

**(7) 工作流与项目管理**

- 1) 任务流程。
- 2) 任务的分类管理。
- 3) 任务统计的图形化处理。
- 4) 任务处理。
- 5) 消息系统。
- 6) 项目维护。
- 7) 计划管理。
- 8) 计划跟踪与控制。
- 9) 团队沟通与协同。
- 10) 项目工时与成本。
- 11) 客户管理与沟通。
- 12) 软件开发过程质量管理。

**(8) 工程变更管理**

- 1) 工程更改的输入。
  - 2) 功能更改审核生效。
  - 3) 工程更改单的管理。
  - 4) 工程更改与技术文件的触发互动。
  - 5) 工程更改的历史查询。
- (9) 用户与权限管理**
- 1) 构建企业组织结构。
  - 2) 维护 PDM 用户。
  - 3) 规定用户基本使用权和物料级的数据权限控制。

制。

- 4) 支持按照组或角色方式进行授权。

**(10) 日志管理**

- 1) 完整记录用户所有操作。
- 2) 自定义查询用户对系统的关键操作情况。
- 3) 筛选查询用户操作情况。
- 4) 管理、维护用户时间记录。

**(11) 系统定制管理**

- 1) 定制各种数据对象的属性定义。
- 2) 定制各种数据对象属性在特定界面的显示方式。
- 3) 定制各种数据对象属性的界面编辑形式。
- 4) 定制物料对象在 BOM 树上的显示属性。
- 5) 定制数据查询结果的输出格式等。

**(12) 集成与接口**

1) ERP 接口：包括产品结构输出、物料主数据输出和工艺信息输出。

2) AutoCAD 接口：管理环境批处理 DWG 文件，自动提取物料清单，快速构造产品结构；在设计环境直接存取产品零部件信息、BOM 结构和图形数据。

3) 3D-CAD 接口：在设计环境直接存取产品零部件信息、BOM 结构和图形数据。

4) CAPP 接口：在工艺设计环境直接存取产品零部件信息、BOM 结构和工艺图形数据。提供可配置的与 SIPM/CAPP 集成接口。

**1.2 软件特点**

SIPM/PDM 具有如下功能特点：

1) 全面管理产品全生命周期所有技术数据。

2) 真正分布式管理，架构可伸缩，既适合中小型企业的产品数据管理，又适用于大型集团企业的异地远程开发协同。

3) BOM 引擎化和视图化，设计、工艺、生产、采购、装箱等业务环节各取所需数据均可满足。

4) 设计/工艺/制造管理一体化，最大限度支持业务流程的重组与优化。

5) 完善的设计更改控制和记录，设计与变更一体化管理，完整记录设计演化历史，并提供追溯功能。

6) 提供产品的设计和变更的试制、生产准备、投产等全过程控制。

7) 扩展供应商、成本、采购相关管理功能，使产品开发相关业务职能更加完善和集成一体化。

8) 提供派生设计功能，并记录派生关系，为产品改型设计提供最大便利。

9) 创立零部件的通用度评价体系和派生关系记

录,为企业设计标准化提供了评估依据。

10) 人性化集成界面设计,所有业务过程可以通过统一界面完成。

## 2 iMAN

### 2.1 概述

EDS公司的Information Manager (iMAN)是专门为整个企业提供产品数据管理的PDM产品。iMAN的6版本正式改名为iMAN。iMAN是建立在Web浏览器技术平台上,充分利用国际互联网的巨大功能,为全球范围的集团级企业服务的PDM产品。

iMAN系统由数据仓库与文档管理、工作流程管理、产品结构管理、扫描图像管理和系统管理等五部分组成。这些功能模块可帮助用户管理和维护产品数据。系统还提供集成工具包(ITK),帮助用户把第三方的应用系统集成或封装到iMAN的系统中。在iMAN系统中,还可通过设置环境变量或用ITK开发一些专用模块的方法,使该PDM系统能够满足用户的特殊要求。

iMAN与CAD/CAM/CAE中的UG及solid Edge两个系统具有紧密集成的关系。作为iMAN的一个子集,UG/Manager将有效地管理UG的装配件、分析数据与文档管理。

### 2.2 系统概述

#### 2.2.1 iMAN 功能模块

iMAN可用于管理各种各样应用软件产生的文档资料,如CAD/CAM/CAE、MRP-II、文字处理、图像扫描等。iMAN包括如下应用模块。

##### (1) iMAN 基本模块

该模块包括iMAN工作平台、系统管理和封装工具。

工作平台——iMAN的工作平台实际上是一个交互式进行工作的环境,每一个工作平台对应着指定完成的任务,并给用户提组织完成任务的各种数据的工具。通过文件夹、数据集及其他文件管理工具,在工作平台上可以建立各种数据间的联系。

系统管理——系统管理对iMAN的工作环境进行管理和控制。系统管理可以定义一套组织机构。其中包括人员、用户、名称、工作小组、项目人员、各种角色、工具和工作站点等。另外它还需要设置各种访问权限和特权,定义数据仓库的具体位置及大小,构造整个系统的固定参数。

封装工具——iMAN的封装工具使用户可以很快

地将各到应用系统封装到iMAN系统中。用户可以访问和维护这些系统的信息,无需知道和手工查询,激活产生这些信息的原始应用系统。

##### (2) 产品配置模块

PSM是基于配置管理的产品结构管理模式。PSM通过BOM浏览功能进行修改和控制产品的结构,BOM实际上描述了除几何信息以外的产品全部信息。通过CAD几何模型、各种分析结果、NC加工程序和文本说明来描述BOM结构间的关联,从而建立访问工程信息的逻辑通道。在PSM中,选择适当的零部件,在CAD系统中可以构造虚拟的装配。在iMAN中还提供各种有效性管理,如根据日期、系列号、供应商的供应能力等制定不同的规则,便于组织生产。

##### (3) 高级配置模块

在高级配置管理中,PSM还可以提供各种零部件的重用、派生和替换的机制以及根据不同变量条件进行选择不同的BOM表进行比较等功能。在替换中,系统可以通过技术说明显示对于特殊用途需要更换多少个零部件。不同的配置要求可以制定不同的配置规则,对一些可选的零部件进行配置管理。不同的BOM对应产品生命周期不同阶段的定义,比如设计视图和装配视图。通过图示化的BOM对比,用不同色彩显示不同的部分。BOM的比较可以用于建立多层和单层BOM之间在装配或零件级使用上的区别。

##### (4) 工作流程模块

iMAN的工作流程模块提供了过程管理的各种功能,如建立、监督、管理和控制等。在iMAN的工作流程中提供了两种基本的应用流程管理,即审批发放和更改管理。另外,利用ITK开发工具中的企业流程定义模块(EPM)可以制定用户专用的流程控制模式。利用这种自定义的方式,可以很容易地在审批发放过程中,通过计算机建立的或提交一项作业或任务给另外一个独立的部门,进一步开发编辑、审核或批准等工作。在工作流程中,更改管理提供了多级发放步骤,不同的权限审批级别及规定各级审批通过后立即执行相应的有关任务。实际上,更改管理是一个典型的审批过程,其中包括更改的全部过程,如更改申请、执行和发布等。

##### (5) 图像管理模块

iMAN图像管理功能可以管理点阵数据、图文数据、音像数据等。这些数据可以单页或大批量一次性地输入到iMAN系统中。多页的图像数据在iMAN中也作为一个对象(ITEM)进行管理。在图像管理模块中还提供底层的打印/绘图能力。此外,批注(View/



mark-up) 功能促进了并行工程的实施。

#### (6) 开发工具

iMAN 的软件体系是由一组集成开发工具包 (Integration Tool Kit, 简称 ITK) 编写的。换句话说, iMAN 的全部功能模块是建立在这套工具包之上的应用模块。因此, 通过 ITK 可以访问 iMAN 内部的数据, 这就给用户提供了方便的本地化手段, 利用交互方式执行的各项任务都可以通过编程来实现, 从而对用户界面和实用功能进行用户化和本地化工作。

#### (7) iMAN 的接口

iMAN 提供与 UG 紧密集成的接口, 利用这个接口模块, UG 用户可以直接访问 PDM 中的有关数据。同时 UG 中 CAD/CAM/CAE 的数据也可以交给 iMAN 进行管理。iMAN 不仅管理 UG 自身的文件, 甚至可以管理 UG 文件中的内部数据, 如 BOM、零件属性等。无论是在 UG 中, 还是在 iMAN 中修改 BOM 或属性, 两个系统始终保持同步变化。

对 UG 以外的各种 CAD 系统, iMAN 除了提供简单易行的封装工具以外, 还提供了若干种接口, 其中包括 CATIA、AutoCAD、Pro/E 和 Mentor Graphics 等系统。

除了与 CAD 系统集成以外, iMAN 还提供与 ERP 系统的接口, 如 SAP、Oracle 的 ERP 产品, 还提供电子邮件系统 Lotus Notes 以及报告生成系统 ACTUATE 等的接口。

#### (8) iMAN 支持全部硬件平台

iMAN 4 版本和 5 版本可以运行在 UNIX 工作站和 PC 环境下。iMAN-NT 是运行在 NT 操作系统下的 iMAN 系统。该系统可以和 UNIX 的服务器或工作站混合在一起工作。iMAN-NT 中既有服务器也可以是客户机, 其功能和 UNIX 下的 iMAN 一样。iMAN-PC 是在 MS 的 Windows 环境下执行 iMAN 的模块。iMAN-PC 利用 VB 或 VC Basic 开发 PC 客户机上的用户界面。iMAN-PC 利用客户机服务器的体系结构, 充分利用服务器上的服务器功能。尽量减少对 PC 硬件的要求及对网络资源的压力。另外, iMAN-PC/X-窗口的仿真终端模块, 用户利用该仿真终端模块, 在 PC 机上不仅可以访问 iMAN 的全部数据, 还可以管理 PC 上应用程序产生的各种数据, 如 MS 的 Word 文档和 Excel 电子表格文档。

#### (9) Web 技术

iMAN Web 为 iMAN 用户提供一种简单方便的手段, 无需经过培训便可访问企业内部指定的数据。目前可以查阅有关文档和零件数据, 下载各类文件, 浏览 UG 中生成的装配结构图像, 传递各类电子邮件并保存在 iMAN 的邮件中, 今后将逐步提供网上审批、填

表、发放任务, 传送图形等功能。

### 2.2.2 iMAN 数据结构

在 iMAN 数据库中管理的信息都是属于 iMAN 的管理对象。在 iMAN 中包括以下几种管理对象: item、数据集、表格和文件夹。一个 iMAN 的 item 可以代表一个产品、部件或零件等。iMAN 的数据集提供了管理各类应用软件所产生的文件的手段。例如, 技术说明文件或 CAD/CAM/CAE 系统产生的图形数据文件, 都可以用原来的点阵形式或其他任何计算机形式的文件存放在 iMAN 系统中。iMAN 的电子仓库 (volume) 可以看作是一个物理上的硬盘, 它通过 NFS 方式和用户联系在一起。iMAN 数据集一方面指向这些文件存在的电子仓库; 另一方面还指明对这些文件进行操作的工具。iMAN 的表格数据将各种属性信息直接存入关系型数据库中。iMAN 提供了一套预先定义的属性和用户自定义属性的手段。文件夹是 iMAN 用来建立数据间联系的一种方法。例如, 可以建立一个文件夹, 其中包括使用某一种材料的全部 item 的零件。iMAN 还可以给 item 规定一些预先定义的关联性质来管理与该 item 有关的信息。通常使用的三种关联性质包括: 必须的、专用的和说明的。其他关联性质也可以自行在系统中加以定义。

iMAN 的基本配置包括一个位于服务器上的 iMAN 数据库、位于文件服务器上的 iMAN 文件柜以及连接在网络上的全部 iMAN 用户的工作站。iMAN 的全部特征数据 (meta-data) 存放在数据库中, 而实际的文件则存放在 iMAN 文件柜 (或称电子资料室) 内, 所有的用户都可以访问这些特征数据和相应的文件。

许多公司将 iMAN 应用在不同的地区, 管理着多种多样的应用程序。在每个地区有成百上千个用户。为了优化网络的性能, 在每个地区都有一个 iMAN 数据库和多个文件柜。各地区的全部数据 (包括特征数据 meta-data 和文件) 都存放在本地服务器上。当各地区间需要共享信息时, 本地服务器将输出 (export) 这些共享数据的目录信息, 并存到公共的共享目录管理服务中。其他地区的用户从共享目录管理服务器上查到所需文件后, 再从指定的地区将该数据拷贝到本地服务器上。每个复制的 item 中, 不仅保留原来的信息, 同时还包括该信息原属哪个地区的, 只有属于本地区的信息, 用户才有权对它进行修改, 其他地区接收到的只是只读的数据拷贝。当一个 item 需要由其他工作小组进行修改时, 可以把该数据的所有权转移到另一个地区。

### 2.2.3 iMAN 管理对象的类型

iMAN 管理的是产品及其整个生命周期内的全部数据,如技术说明、标准规范、工程图样、加工程序等,还包括生产过程的文件、更改单、用户使用手册、维护说明等技术文档。凡是在操作系统下可以生成或修改文件的应用程序都可以封装到 iMAN 进行管理。大的、复杂的文件如 CAD 和描程序产生的图形、图像文件;小的如 MS 的 Excel、Word 文档或简单的 ASCII 文本文件,都可用 iMAN 进行管理。

iMAN 还可以管理存储在文件中的数据,iMAN/UG 接口模块就是其中一个典型例子。UG 零件中的属性数据可以由 iMAN 来管理,无论用户是在 UG 还是在 iMAN 中修改这些属性参数,在 UG 和 iMAN 两个系统中,始终保持同步变化。UGS 公司已经建立了四十多种应用程序的集成接口。

### 2.2.4 安全保密

在 iMAN 中对数据访问及使用的安全保密分成不同的控制级别,如数据、功能,用户、工作小组及整个企业等。访问控制表(Access Control List,缩写成 ACL)规定了用户及其工作环境中对 iMAN 的管理对象进行访问的权限。对 iMAN 管理对象进行安全保密控制的模式为:所有者——工作小组——岗位(或称为角色)——系统——全部等五级。对某一个用户或用户工作小组,可以指定他访问 iMAN 中每一个或某一类管理对象的读写权限。iMAN 还允许每个管理对象的所有者自己控制对它进行访问的权限。在 iMAN 中还可以规定哪些角色可以修改哪些管理对象。同时,一旦设计过程完成以后,相应的数据和文件可以立即存档和发放。

在 iMAN 控制下的系统级的文件不受操作系统提供的保护权限的限制。用户在 iMAN 环境下,可以选择把这些系统级文件在 iMAN 的环境中隐藏起来,在 iMAN 中保存的文件可以用加密算法,将其真实名称替换后存放在操作系统内;也可以选择不用这种加密方法,在操作系统下很容易找到相应的文件。

除了数据以外,iMAN 的菜单也是可选择的。利用菜单入口隐藏(MES)功能,可以把不同用户使用的下拉式菜单有选择地隐藏起来。

## 2.3 用户功能

### 2.3.1 数据仓库及文档管理

iMAN 管理功能可保证产品数据的集中管理和安全性,提供多用户的并行访问和防止无权人员的修改。

在设计过程中,用 item 的不同版本来描述数据集的不同设计变化,在 iMAN 控制下的全部文件得到 iMAN 的保护和控制。当一个文件(普通文档、CAD 零件等)发生变化时,iMAN 自动将原先的版本保留,同时产生一个新的版本。每一个 item 版本包含了定义该 item 所需的全部信息,其中包括数据集(即文件)和属性(meta 数据)。一旦 item 对应的表格与表格的内容,或对应的功能发生变化,则自动产生 item 的新版本。该新版本的指针首先是对准原先版本的信息,直到数据集成属性确实修改后才指向新版本。因此,同一个数据不用作多重拷贝,这样可以大大减少对数据库和磁盘空间的需求。

为了防止两个用户同时修改同一个数据,iMAN 采用 Check-in 和 Check-out 机制。当一个用户打开某个数据进行编辑时,iMAN 自动对该数据“写”访问权进行加锁控制,任何其他人都不能对它进行修改。只有当原先打开该数据进行编辑的用户将该数据重新放回数据库或关闭该数据时,该数据才恢复原来状态,即允许其他授权用户进行修改。手工的 check-Out 命令可以对数据设置一个长时间的锁,防止该数据发生变化;Check-in 则用来解除原先设置的锁使数据恢复到可修改的状态。

当一个用户被分配到另一个项目后,原来 Check-out 的文件,无须用 Check-in 解锁便可转给另一个用户。用户可以把自已的名单附在某一个对象的说明列表上。在这张列表上可以看出该在什么时候 Check-out 应和 Check-in。此外,用户还可以查询全部 Check-out 的或某一个事物被锁的整个历史。

用户可根据自己的权限访问数据信息,不同的保护级别提供给不同的访问人员、工作小组或角色。当一个设计版本完成后,可以将它存起来防止被修改。iMAN 电子仓库是一个逻辑仓库,当一个文件存入该仓库后,它的状态被设定为不能修改的 item。

iMAN 在审批与发放过程中,还提供了一种方法来规定不同权限的变化过程。CR(审批流程)是 iMAN 的应用模块,它允许有多级的审批过程。对于事先规定文件的发放过程,如审核、批注、电子签字等过程均由计算机自动地进行管理。

iMAN 系统管理员根据对象的类型和有关的工作人员,预先编制一套审批发放模型。这个审批发放程序包括审批过程、级别以及每一级中电子签字的角色和审批结束阶段。在一个审批过程中,可以有任意多级审批,在每一级中可以指定一个或多个角色进行审批,在同一级还要制定表决的规则。每个人在电子签字时有三种选择:同意、反对、弃权。批注可以和每一种表决结果附在一起。

### 2.3.2 工作流程和流程管理

iMAN 工作流程模块可对企业级的工作流程进行管理。它可以创建、执行、监督和控制工作流程。用户可以建立串行或并行的工作流程,使得工程技术人员并行地协调工作。当用户打开自己的收件箱时,就可以看到最新的工作任务清单。项目发起人或负责人可以随时查看整个项目的执行状态,直到该工作全部完成。iMAN 的工作流程管理提供 CR (审批流程) 和 CM (更改管理流程) 两个应用模块。这两个应用模块可造成专用的工作流程。工作流程可以与指定工作小组和专门的事项相关联,同时还有专门的日志文件记录该流程中的全部管理信息。

CR 应用模块可以定义文档的审批过程,它提供一套很方便的工具来定义审批的级数和批准的人数。工程更改单也可以用 CR 程序来管理。CM 描述了典型的更改管理,其中包括提交、执行和发布更改的各项步骤。CM 利用一套表格来归纳和审批提交更改的数据。这些更改的数据既可以是一个数据集,也可以是一个 item,或是一组对象(object)。整个更改流程包括 8 项任务:更改申请、批准申请、权限影响分析、实际影响分析、权限更改指令、实施更改指令、批准更改指令及有关说明。在这 8 项任务中,实施更改指令,激活 CR 审批流程,对实际发生的更改数据进行审批。随着更改管理程序的一步步执行,自动激发下一步任务,并给相应的用户发送信息,保证整个过程连续工作。

在整个更改管理过程中可以指定审核者和批准者。审核者可以对修改作批注,但没有批准的特权。只有批准者有权同意或反对这项修改。iMAN 的工作流程管理还给用户提供了组织专门流程的框架。在所有的 iMAN 工作流程中,可以设置专门的规则,在每一个阶段还可以执行其他过程或任务。自定义的工作流程可以用工作流程工具包来定义,该工具包称为企业过程建模 (EPM)。它包括了“C”语言的应用程序接口 (API) 和动态连接库 (DLL)。用 EPM 定义的工作流程可以按指定的步骤对专门的数据修改有关的状态或者根据先前的操作进行流程管理。在 CR 和 CM 中还可以在什么时候、任何地点执行用户自己编写的外部执行程序。

### 2.3.3 产品结构管理

iMAN 的产品结构管理 (PSM) 提供管理产品的结构定义和配置,图形界面使用户可以很容易地浏览和维护产品的结构 (装配、子装配和零部件),在图形界面上,使用标准的工具,如复制、剪切和粘贴来维护产

品的结构。当然,iMAN/PSM 定义了产品的 BOM,对应于产品生命周期中不同阶段,如“设计阶段”,“组装阶段”等。不同的 BOM 代表同一个产品的不同角度所定义的信息。

iMAN/PSM 把产品定义的全部数据,包括几何信息、分析结果、技术说明、工艺文件等都和产品结构建立起联系。iMAN/PSM 提供了各种工具去查询 item 和访问有关的数据。“引用查询”功能使用户能够很方便地得到某一项变化所造成的影响,即查询原来使用变化前的 item 的全部结构。iMAN/PSM 提供的标准功能使查找大型 BOM 时,可以很容易找到指定结构的位置。产品结构配置的原则是由用户自行选择的,可以根据 item 的版本或工作状态来决定。iMAN/PSM 还支持有效性管理。例如,根据日期或系列号或供应商的供应情况来配置,从而确保生产制造的正常进行。

产品结构编辑器 (PSM) 为用户创造和维护 BOM 提供了强大的工具。每个 BOM 用一个锯齿形体系结构表来表示,用户可以选择水平或垂直两种显示方式。在显示 BOM 结构树时,用户有两种方法控制显示的层数:

- 1) 在 BOM 中的装配件可以全部显示内部的零部件,也可以全部隐藏掉。
- 2) 只详细显示指定零件的结构数据,其余都隐藏掉。

iMAN PSM 支持精确和非精确两种配置。精确配置是针对确定版本的 item 的固定结构,非精确配置代表 item 的动态结构,它可以是若干个版本的组合。根据指定版本进行配置时,每一个 item 根据 CR 规定的状态来确定;根据日期有效性或系列号有效规则来配置产品时,可输出指定的 BOM 表供用户编辑和查阅。用户还可以指定输出一个包括所有版本的全部 item 的清单,从而对样机进行修改,生成新的产品。

在 BOM 上选择一个 item,便可立即打开一个工作窗口,用户在这个窗口内可以看到与该 item 有关的全部信息。例如,零件文件、技术说明等。当使用 iMAN/UG 集成模块时,在 BOM 中有多个 item 可以选择。如果选择 UG 二维 CAD 模型的虚拟装配,则在窗口中自动显示它的二维模型。当在 UG 的装配模块内或在 PSM 中修改装配结构时,则另一面会自动修改。

iMAN 的高级 PSM 还支持设计重用和派生,并能解决替换变量化、多视图及 BOM 的比较等功能。在产品结构中,为了特殊要求,允许作适当的替换,并附带必要的说明,显示内部变化的个数。例如,生产车间可以根据库存数量或采购价格来决定采用哪一个 item。

变量化产品配置是根据可选项的定义和产品配置规则来进行匹配的。例如,在设计汽车仪表盘时,同一个汽车型号有各种各样立体声收音机的型号可供选择,不同的汽车可以根据自己的规则来选择不同的立体声收音机。如果对仪表盘需要作修改,设计人员可以规定全部变量依然有效。反过来,针对某个汽车型号的BOM,根据变量定义规则,将提供一个精确的配置方案。

多视图功能为不同部门从不同的角度来配置BOM提供依据。例如,制造部门在“制造视图”中需要包含工艺信息,并把这些信息传到ERP环境中。针对产品、装配等可以定义多层视图的BOM,多层视图功能并非需要高级配置管理的技巧,从任何一个BOM的视图中均可访问产品的全部数据。

#### 2.3.4 分类与追溯

iMAN提供零部件分类的功能。在iMAN表格内存储了一套属性。根据这些属性可查到类似的对象,这相当于提供了外部分类的界面。

### 2.4 实用功能

#### 2.4.1 通信和通知

iMAN的电子邮件可以发布消息和数据。iMAN邮件把数据分发给其他iMAN用户是通过指针指向对应的数据,数据本身没有多重性。当iMAN向一个用户发送邮件信息时,操作系统也发出一个iMAN的邮件。iMAN的工作流程就是利用iMAN的邮件系统和操作系统的电子邮件给各个用户发送重要事件和必须的通知等信息。

在新版本iMAN中还提供与各种公共的电子邮件管理系统的通道。利用这种通道,iMAN的用户可以根据数据管理范围,直接把数据通过e-mail的方式传送到Internet或Lotus Notes中。

#### 2.4.2 数据传递

iMAN始终监视着各种数据,它随时掌握用户需要些什么数据,在什么地方和使用什么应用软件来处理这些数据。iMAN将自动地把数据从一个地方搬到需要审批的那个地方,根据用户共享数据的要求,一般只有数据参数,而真实的数据依然存放在数据仓库内。所有的其他引用都只是—个指向该数据的指针。输入和输出数据的功能仅用于两个不同的iMAN数据库之间。

#### 2.4.3 数据转换

iMAN STEP提供专门的数据转换接口。通过

STEP AP203接口可以实现不同PDM之间的数据交换。STEP AP214提供不同部门,采用不同CAD系统的PDM之间数据交换。

#### 2.4.4 多媒体数据服务功能

iMAN多媒体服务提供了管理点阵和矢量数据的能力,并能管理混合的包括声音和图像等在内的数据。图像数据可以单个也可以大块输入,多页图像可以作为一个item放在iMAN进行管理。图像可以用于审核、修改、打印、存储和发放,声音文件可以作为邮件管理,快速发送到指定的用户。其他第三方文档管理系统可以通过封装或集成的方式进行管理。

UGS提供的ProductVision(PV)软件,正是满足各级管理部门在iMAN环境内,查阅加批注CAD图形数据的要求。通过PV(以将设计制造供应商和最终用户联系在一起,共享产品的主体模型或工程图纸。

#### 2.4.5 iMAN产品结构管理

产品结构是由对象间的关系组成的,在整个产品生命周期内,可建立和管理产品的定义和结构,提供产品结构的多视图。产品结构和配置方面向用户提供图示化的界面,使用户能对产品结构进行创建、浏览、编辑。产品结构用树状结构表示。

通过建立产品item而建立BOM,将各个子系统item用指针指向统一的BOM,从而建立一棵产品树。iMAN的产品结构管理用户提供了浏览、检索、编辑、多视图管理的功能,为产品数据或文档的变更提供了自动传播和动态跟踪,并且支持产品数据版本管理。根据用户的要求,可以快速地配置产品结构,即产品配方管理。iMAN提供了产品结构BOM表的生成功能,实现与ERP系统的联接。

在产品结构与配置管理中,主要提供以下功能:

- 1) 创建产品的定义和结构。
  - 2) 产品结构树的多视图管理。
  - 3) 浏览和查询产品结构树。
  - 4) 编辑产品结构树。
  - 5) 为产品数据或文档的变更提供自动传播与动态追踪功能。
  - 6) 支持产品数据的版本管理。
  - 7) 根据用户的要求设定配置规则,用户可以根据不同的情况得到产品的适当配置。
  - 8) 在产品结构方面,提供BOM表的生成功能,并通过BOM表预留将来与ERP的信息接口。
  - 9) 实现人员模型中各角色对产品结构的要求。
- 对于大型BOM表之间的比较,完全可以在几台

工作站上同时进行数据比较,最后达成整个 BOM 表的比较结果,iMAN 可以使 BOM 表的比较能在硬件所能够达到的最佳性能下进行。

#### 2.4.6 iMAN 分布式数据处理

在分布式数据处理中,着重解决的是如何处理信息共享和系统效率之间矛盾的问题。目前,所谓全分布式的数据处理技术本身已经较为成熟,但在工程领域,由于数据量大,数据之间的关系复杂,所以系统效率成为极为突出的瓶颈问题。因此不能简单地套用商用数据库管理系统的分布式处理方法,必须找到一个既能解决工程领域的信息共享要求,又能保证系统具有较高效率的解决方案。在这方面,iMAN 可提供独到的处理方法。

在分布式数据处理过程中,iMAN 既能解决异地数据共享的问题,同时也解决了一般分布式数据管理系统中效率太低的问题。在 iMAN 中数据采用主动发布的形式 (publish),区分了非共享型数据与共享数据。在处理方面,对不需要共享的数据不提供分布服务,使得共享数据能得到较高的响应速度。在数据共享时,提供不同策略,当用户仅仅是查看产品结构时,iMAN 并不移动所有的产品信息,而只是将类似于目录的描述信息提供给用户。只有当确实需要对数据进行操作时,才执行数据的移动,这些措施进一步保证了系统的效率。同时,这样的体制也有利于进行权限控制,符合系统分层管理的需求。

#### 2.4.7 备份与归档

iMAN 具有专门的备份与归档功能,其中包括三种类型的备份和恢复工具:包含 Oracle 的备份、iMAN 自身备份和有选择的备份。包含 Oracle 的备份是最合适的备份,并尽可能少地占用系统暂停的时间,它利用 Oracle 的功能将数据库中的数据备份到文件中,并保存在 iMAN 的环境中。iMAN 自身备份是周期性的备份功能,可以每三或四周进行有选择地备份,实际上可以只备份数据库中的数据或 iMAN 文件柜中的文件。例如:可以专门备份某个项目的数据或经常使用的数据。为了进一步保护 iMAN 文件柜,iMAN 还支持镜像功能。在最后一次备份以后存入的新文件或更改过的文件或光盘均可同时复制到另一个文件柜内,一旦原来的文件柜受到破坏,就可以从那个文件柜中取回这些曾备份的新文件,对于长期不用的数据,iMAN 可以用归档功能把数据存到磁带或光盘上,以便今后使用时再恢复回来。

### 3 Windchill

#### 3.1 Windchill Info \* Engine 数据页

过去已经限定在传统和当前信息系统中的信息和过程,现在可以使用 Windchill Info\* Engine 将它们无缝地和逻辑地连接起来为供应链中的每一个成员建立集成的企业解决方案。使用 Windchill Info\* Engine 的信息采集和 Web 支持功能,广义企业的成员可以在需要时获取他们所需的信息。

##### (1) 基于 Web 的体系结构

Windchill Info\* Engine 把 Web 用作一种通用平台来提供一种可伸缩的应用框架,将松散耦合式系统中的信息汇集到功能强大的业务过程解决方案中。Windchill Info\* Engine 使用 Web 浏览器可以快速生成和交付具有特殊作用的应用,为终端用户提供他们所需的准确信息。Windchill Info\* Engine 可以使用现有的企业内部网、企业外部网或 Internet 进行通讯,它能够与产品和过程生命周期有关的每个成员提供信息。

##### (2) 顶向下的集成方法

Windchill Info\* Engine 提供一种自顶向下的应用集成方法,它并不依赖源代码编程,而是靠组合多个信息源来完成应用集成的。它之所以具有这种能力在某种程度上是因为它使用了工业标准的 IT 基础结构、网络协议和应用编程接口 (API) 等功能,已比如结构化查询语言 (SQL)、开放式数据库连接 (ODBC)、Java 数据库连接 (JDBC)、微型目录访问协议 (LDAP) 以及超文本链接标示语言 (HTML) 等。这种方法能让商业分析家从业务过程的角度开发出综合解决方案。

##### (3) 综合应用框架

Windchill Info\* Engine 由多个模块组成,提供了一种完善、开放、可扩展的平台,企业可以在这个平台上建立一种综合产品生命周期信息应用。它产生的商业解决方案充分利用了现有技术、硬件和传统系统从而利用了企业过去的大部分投资。Windchill Info\* Engine 建立了一种无边界环境,在规定的参数范围内,它能让用户:

1) 查找、选择和获取分散在网络上的信息资源。用户不需要知道这些信息的技术特性、保存这些信息的原始服务器、或者提高数据传递效率的网络技术即可访问信息。

2) 使用 Web 浏览器查看基于网络的信息,信息用户不必用很多种应用软件配置桌面客户机即可访问各种资源。

3) 检查集成的异构信息。这些信息从企业资源中

提取而来它们被放进企业每天的业务过程中,从而把业务过程重组概念转化为实际的实现方法。

#### (4) 支持消息系统和事务系统处理方法的服务器

Windchill Info\*Engine 服务器从根本上支持消息系统和事务系统的处理方法,提供功能强大的提取,变换和装载(ETL)功能,使用它提供的创建、阅读、更新和删除等功能,它可以支持双向信息访问,并支持从几乎所有系统中提取信息。另外,它还能提供分类,连接、合并和交叉等数据处理功能,以进行数据变换和格式化操作。Windchill Info\*Engine 服务器有两个主要组成部分:

1) Webject 库。Webject 库能让用户自动把数据库模式和对象模型映射到部门应用中。Webject 标签是一些能被插入到 HTML(超文本链接标示语言)模板文件中的 HTML 对象。当 Web 服务器在 HTML 文件中遇到 Webject 功能标签时,对象标签被解析为执行某条指定的命令。之后 Windchill Info\*Engine 服务器对目标系统执行了一个动作。Webject 可以用来完成许多动作,其中包括以下几种:建立连接应用的通用 API(编程应用接口)、定义显示格式以及描述任何一个应用外部的计算操作等。

2) 模板库。模板库定义特殊执行过程。通过 Windchill Info\*Engine 接口,模板过程为用户提供了访问一般应用和相关通用的显示屏幕功能。模板子单元定义材料清单(BOM)等对象之间的相互关系,并定义创建零件等任务之间的相互关系。Windchill Info\*Engine 模板提供了一个柔性机制、用来确定综合应用必须访问的信息,以及信息显示的模式。模板库还提供了样品模板,它可以简单地定制或扩展,以满足特殊的业务要求。

### 3.2 Windchill ProductView 数据页

ProductView 不仅用于提高产品设计及开发小组的生产率和通讯能力,而且用于提高整个广义企业的生产率和通讯。它不但涉及设计、销售、支持、会计、质量保证、制造和装配,而且还涉及供应商和客户。ProductView 有一个直观用户界面,该界面极少需要培训,甚至不需要培训,用户可以把它定制成只包含完成他们工作所需的控制。为了理解设计可以使用各种显示选项查看模型,可以测量尺寸、创建截面以及计算质量特性。

#### (1) 异构产品信息查看

ProductView 能让整个企业的用户查看许多种与产品有关的信息——从诸如微软 office 文件这样的文件到 2D 图纸和 3D 模型——所有这些信息都是在一种基于 Web 的常用框架中。用户不必知道文档保存在

什么地方,就可以在一种单一的集成环境中浏览系统、查看模型和导航产品结构及相关信息。为了建立支持工程变更请求的材料,并支持在线概念审查用户可以查看、询问和标记各种 2D 和 3D 标准及自然格式。

#### (2) 产品结构和元数据查看

开发和制造一件产品需要很多类型的数据。描述一件产品的数据称为元数据。产品结构提供了以一种容易理解的方法查看信息。Productview 提供一种导航产品结构的直观方法。相关信息可以是文档、图像、电子数据表,URLs 或列表等形式。元数据直接与容易查看的组件关联。

ProductView 提供了查看 Windchill、Pro/PDM、Pro/INTRALINK、optegra、Sherpa 管理的产品结构信息和元数据,以及 CATIA 和 CADDS5 文件系统的功能。产品结构可能包含各种各样的数据类型,其中有三维模型、二维图样、办公文档、图像、影片和 Division MockUp 仿真等。利用包含不同文档、部件或组件的窗口,可以简单地查看到整个部件,单一组件或任何组件的组合。

#### (3) 3D 模型查看和分析

ProductView 提供独立于 CAD 的基本查看功能,它能让产品和过程生命周期中的参与者计算质量特性、建立效果图像、通过拖曳对象移动或旋转它们来剖析部件、穿透大型部件以及建立部件的截面等。通过使用穿透模式,可以达到快速调查设计背景的目的,穿透所有部件设计是研究部件设计的一种普通方法,透视查看还可以被用来为设计提供在现实中看起来象什么的更准确的表示。

ProductView 包含动态切片法等高级分析工具,而能详细审查部件设计。在 2D 维和 3D 模式中,用简单的“拖-放”鼠标操作进行动态定位,可以调查各种结构实体模型。设计的不同配置可以被保存起来以便以后再使用。

ProductView 测量工具能测量准确的长度、角度、半径和曲面间隙尺寸,还能测量点、中心、边缘和曲面之间的距离。另外,还可以查看源 CAD 系统的特性和属性信息。ProductView 三维模型查看和分析工具能向非 CAD 用户提供宝贵的产品信息,从而支持企业协作。

#### (4) 2D 查看

ProductView 支持大量 2D 图样、矢量、图像和文档格式,从而能让企业参与者不需要访问源应用,就能查看、处理和覆盖几乎所有类型的 2D 数据。另外还提供了 Batch-plotting(批量绘图)工具和定义水印的能力。

#### (5) 关键标记和分析功能

ProductView 能让参与者标记或涂注所有类型的产品信息,并储存和检索这些标记或涂注信息,标记可以作为 HTML 文件通过环球网被发布出去,并由 Windchill 或工作组信息管理系统来管理。例如,工程管理员可以用规定的键设计标准来标记图样,采购人员可以把这一信息作为注释发送给供应商。2D 和 3D 标记工具都包括注释、尺寸,重要部分和标记。同一图样可以作多种标记,并可以查看这些多种标记。另外用户还能够定制注释线宽和字体设置等图形属性。

使用 Division MockUp 可以建立装配、拆卸和操作员交互作用等模拟场景。联合使用 ProductView 标记和分析工具以及基于 Web 的可视化工具,广义企业的成员第一次可以方便地参与产品和过程生命周期的任何阶段,因此,能生产出更高质量的产品,并加快产品上市时间。

#### (6) 通过图形服务器访问产品信息

ProductView 客户机程序可以通过 Web 下载并自动被安装到目标 Web 浏览器中,从而可以方便地部署,以使维护成本较低。有三种客户机程序选项可以选择:

1) 一个 Windows 和 UNIX 本地客户机程序它或者作为独立的程序运行,或者作为 Netscape 或 Internet Explorer 的插件程序运行,这种选择提供最高的运行性能。

2) 一个可下载的、基于 Java 的客户机程序。这一客户机程序提供一个独立于平台的环境它的交互性比 HTML 要强,且不需要在客户机上预安装任何软件就能运行。

3) 一个基本的 HTML 客户机程序。HTML 客户机程序允许在任何平台访问信息,即使像 Windows 3.1 这样只支持一个简单 Web 浏览器的平台也如此。

ProductView 客户机程序通过使用三层客户机服务器体系结构的 ProductView 图形服务器来访问产品信息。这一以 Web 为中心的设计能用一个通用界面访问大量信息源,并能处理许多不同信息格式。ProductView 图形服务器提供了对 CAD、工作组信息管理和文件系统信息的控制访问能力,并可以简单地扩展到支持新的或专用的 CAD、数据库和信息管理系统。通过在开放式应用编程接口 (API) 上访问 CAD 系统图形服务器可以生成准确、独立于边界的表示。这样建立的 3D 几何表示非常紧凑,很容易被分发到网络上。另外,也可以象本地源 CAD 数据一样的拓扑实体描述几何体,因此,精确度不会降低。因为 ProductView 可以成批或快速建立可查看格式并把它们储存起来,以便以后使用。所以现在可以在几秒中内安装和查看一个完整的异构部件。

### 3.3 Windchill Foundation 功能

#### (1) 文档管理

使用了 Windchill 文档管理功能,对企业信息的建立、储存和维护进行管理,就象选择一个超级链接一样简单。对多用途网际邮件扩充协议 (MIME) 的标准支持,可以让用户把对不同文件类型的查看功能和文档编辑功能,编成简单的 Web 浏览器任务,文档管理功能包括:

1) 一个数据仓库,它能够通过控制权限,来保护信息不会被未授权用户修改。

2) 登录和退出功能,用于把文档提交给数据仓库以及从数据仓库中提取文档。

3) 版本控制和历史记录功能,用来跟踪数据更改、记录根据已有文档建立的新文档的来源,并能浏览历史数据。

4) 全文检索功能,Verity Search'97 引擎支持这一功能,无论文档的格式是什么,它都能提供快速确定和访问内部及外部产品数据的功能。

#### (2) 结构管理

Windchill 结构管理功能通过定义零件间的关系,建立一种单层结构。层次结构可以描述系统、子系统以及部件的整个产品细节。使用简单的 HTML 导航功能用户可以访问材料清单 (BOM) 的一种功能视图,它会随用户在企业中角色的不同而改变。比如一名工程师可以操纵设计视图,而一名装配规划人员则只能在规划视图中工作。

#### (3) 生命周期管理

Windchill Foundation 使用生命周期管理工具,管理信息随时间而完善的方式。一个 Windchill 生命周期包含了一系列连续的阶段和阈值,它们能确定某个对象当前的状态。并能确定该对象进入下一阶段所要求的阈值条件,通过把工作流程与生命周期阶段和阈值联系在一起,生命周期管理的对象可以在它们的生命周期状态间被自动移动。这种由工作流程支持提供的生命周期的自动转变,可以使企业获得更高的生产率。生命周期管理功能包括:

1) 已有的生命周期模板,可以修改和重用它们,以快速建立一种 Windchill 环境。

2) 生命周期定义模板,它们可以让用户快速定义对象在完善过程中所经历的阶段和阈值。

3) 生命周期管理器模板,它们能提供对所有过程模板的访问,并支持对这些模板所采取的操作如:建立、更新、查看、重命名、删除、登录,退出、导入以及导出。

4) 基于状态的访问控制,它能根据对象的状态,



米为用户或组授予访问权限。

5) 基于项目的角色定义,它能定义某个特定项目中包含的角色——这样就能简化生命周期模板的重要工作,另外它还能把工作人员分配给每个角色。

6) 基于事件的通知,它能保证随着对象在生命周期状态间的变化,有关对象的重要信息可以及时送到适当的工作人员手中。

#### (4) 工作流程管理

通过提供一种统一而又灵活的过程管理框架, Windchill 工作流程管理功能支持产品和过程的差异,这些差异可以提高企业在当今市场上的竞争优势。Windchill 工作流程管理功能能让用户积极指导和监视他们独特的商业过程——帮助他们制造出优秀的产品,并能缩短上市时间,降低开发费用。该功能中还包含了工作流程模板示例,可用来快速裁剪和实施通用 Windchill 活动。

#### (5) 工作流程定义

Windchill 图形化过程编辑器能快速简单地定义工作流程。过程开发人员能建立工作流程模板,并能对它们的版本进行控制。为了定义一系列复杂的工作流程, Windchill 提供了一些预先定义好的活动组成部分、一个集成的 JAVA 编译器,并提供对以下几方面的支持:嵌套过程、过程引用、分支、特别活动的建立、表决。过程与任一 Windchill 对象间的同步、定时器、合并、重复次数、以及基于响应和属性的路由。

#### (6) 工作流程参与

工作流程初始化过程建立了工作项目的 URL。当打开某个工作项目的 URL 时,工作流程系统会生成 HTML 表单,这些表单中包含了与参与人员交互的响应选项和可变域。任务会被自动分发到参与人员的工作列表和电子邮件中。主管可以重新分配任务,以平衡工作负荷或消除瓶颈,工作流程系统还提供了活动表单,可以方便地对它们进行动态定制,以便能为每一个活动提供信息,可以修改参加人员,并且参加人员并不限于某个当地机构。相反,供应商、合作伙伴以及其它企业都可参与这些自动化过程,从而提高了协作的水平。

#### (7) 工作流程的监视和报告

工作流程管理员可以用图形的方式,监视和管理正在运行的工作流程,并能审查已经完成的或人为中止的过程。工作流程和活动的详细信息可以指明某个活动当前的状态、它什么时候开始、什么时候结束、它的时间段、参与人员以及变化值。可以加快、跳过、暂停和恢复、或者中止某些过程和活动。

#### (8) 系统管理

Windchill Foundation 系统管理功能包括:

1) 控制访问。Windchill Foundation 可以定义、实施和管理灵活的访问控制策略,从而确保了在一个项目的生命周期特定阶段中,只有适当的用户和组才有访问某个域中对象的权力。比如,可以建立一种规则,规定在对象评估阶段,只有工程师才有权限修改某个对象的需求。

2) 联邦式管理。Windchill Foundation 为管理员提供了一些工具,用来配置和管理从 Windchill 中访问企业信息资源的权限,这些工具包括 Windchill 和 PDM 系统、企业目录服务以及传统数据库。

3) 联邦式目录服务。由公司自己目录服务(如:支持 LDAP 的企业目录)维护的用户和组信息,可以被调入到系统并能与 Windchill 相关联。比如当某个用户的属性在 LDAP 目录中发生变化时, Windchill 会自动使用修改后的值。这一功能使企业可以充分利用已有的基础结构,并能确保 Windchill 中的信息始终是最新的信息。另外,使用联邦式目录服务公司可以把 Windchill 加速推向新企业。

4) 域定义。Windchill Foundation 能协助企业组织信息。通过设置域,企业可以把使用相同管理策略(如访问控制、索引和事件声明)的对象,组合到一起。比如一家正在安装一种帮助桌面应用的企业可以使用通用 Web 搜索技术来识别信息。通过基于事件的策略,以控制某个对象应该在什么时候进索引的集合中,什么时候移出, Windchill 索引策略定义能自动完成维护索引这种繁琐的工作。

5) 项目和角色定义。对商业对象的组织,它能将预先定义和公司定义的项目角色,映射到特定的用户和组中。

### 3.4 Windchill 工作小组管理器

#### (1) 关联产品结构管理

Windchill 工作小组管理器能让用户建立、处理和定义 CAD、ECAD 等工程设计工具或软件与 Windchill 产品结构之间的关联。可以在产品结构中建立和管理不同配置类型的关联以反映不同的设计要求和标准。通过把相关设计信息集成到企业的业务系统中, Windchill 工作小组管理器能保证广义企业中需要设计信息的员工能访问到所有信息。销售和市场部门的非工程师第一次可以直接参与产品开发过程。

#### (2) 灵活的设计过程

Windchill 工作小组管理器既允许采用自顶向下的设计方法又允许采用自底向上的设计方法,因此,它能保证在设计过程中具有最大的灵活性。自顶向下的设计方法使设计人员在设计产品时具有“整体思想”,而自底向上设计方法则允许设计人员通过一次增加一



个组件来设计产品, Windchill 工作小组管理器能控制任何工程设计工具部件结构和 Windchill 产品结构之间的双向关联变更。

### (3) 充分利用产品结构信息

因为使用了标准的 Windchill 产品结构, 工作小组管理器起着管道作用, 该管道为广义企业提供产品结构信息。Windchill 产品结构不仅融汇了来自其它 Windchill Factor 系列应用的信息, 而且融汇了企业资源规划 (ERP) 等其他企业应用的信息。这能让企业从他们现有的 IT 投资中获得最大的利润

### (4) 可访问的数据库

Windchill 工作小组管理器在逻辑上保存了 windchill Foundation 时期的 CAD 模型、详细图样、技术规范、可视化对象以及替代表示等设计方面的信息。把所有这些产品信息都保存在一个通用的位置, 可以方便需要它的所有人员都能很方便地访问到这些信息。另外 Windchill 工作小组管理器还能控制或管理由工程制作工具产生的模型依赖关系, 从而能让工程师集中于产品革新而不是信息的管理。

### (5) 异构设计环境

Windchill 工作小组管理器能够建立由各种制作系统中的信息组成的产品结构。Windchill 工作小组管理器从根本上支持不同的制作工具 (例如 PRO/ENGINEER, CADDSS), 所以它能使企业集中精力进行产品创新而不用管理 CAD 工具的兼容性。这一强大的功能可以让企业把来自不同供应商、合作伙伴的信息集成到产品设计过程中。

## 3.5 Windchill 网关数据页

### (1) 发布产品信息

Windchill 网关发布机制提取设计工作小组管理工具建立的对象定义、组件数量和产品结构, 来建立 Windchill 对象。除了把信息发布到标准的 Windchill 对象模型外, Windchill 网关还能把信息发布到任何定制模型。在 Windchill 环境中发布的所有对象都有 URL 用来指定外部工作小组系统中的信息。另外, 还会建立 Windchill ProductView 数据并以内容的形式保存起来。Windchill 管理员可以根据发布级别和预定的时间间隔来定义发布准则。这种方法可以灵活控制某个对象应该在什么时候让企业使用, 以及以什么样的频率发布——使用批处理过程隔天发布一次、定期发布或应要求发布。

### (2) 查看产品信息

Productview 能让用户从 windchill 中浏览发布的信息。Windchill URL 代表着源系统中可浏览的信息如 3D 模型或 2D 图纸, 当用户点击它时, 系统会自动

调入 Productvie, 并显示出图像以供研究, 操作和标记。为了提供高速、可伸缩浏览。Productview 数据被储存在缓存中, 用于快速检索。

### (3) 管理发布过程

发布过程是自动的, 管理员可以方便地配置它的行为。系统提供了一套全面的配置选项, 以确保任何类型的信息都可以根据公司的业务需求发布给 Windchill。容易使用的管理工具确保了系统可以建立和执行所有不同类型的行为。

### (4) 安排发布

Windchill 网关提供了一种发布调度机制, 它会指定对象以什么样的频率发布给 windchill。网关按指定的时间间隔搜索源系统, 以找出满足发布条件的对象。Windchill 网关在后台工作, 不需要终端用户的任何干预。根据对象的发布级别来启动发布, 对象的发布级别由管理员在外部工作小组数据管理器中定义。一旦达到了最小的发布级别, 网关会从外部工作小组数据管理器中自动检索最新信息并把它发布给 Windchill。

### (5) 对象映射

Windchill 网关提供了一种机制, 用来把外部工作小组数据管理器中系统定义和用户定义的属性, 映射成 Windchill 基于实例的属性, 管理员可以同时用源系统和 Windchill 检索和显示属性, 之后可以把这些属性以任何方式连接起来, 甚至当对象属性被发布到 Windchill 中时, 可以重新命名它们。

### (6) 定制对象命名

通过使用外部工作小组数据管理器中的任何一种命名方法, 管理员都可以在 Windchill 中给发布对象一种定制的命名模式。如果不存在唯一的识别往 Windchill 对象号码还可以自动叠加。Windchill 网关能发布外部工作小组数据管理器中定义的所有对象类型。默认的对象类型包括 CAD 零件、部件和制图。管理员可以检索已在工作小组数据管理器中定义的对象类型, 并能选择任何文件类型发送给 Windchill。

## 4 CATIA/PDM II 产品开发管理解决方案

CATIA/PDM II 产品开发管理解决方案可用于存储、管理、保护和访问产品数据, 帮助整个企业取得最大程度的成功。

### (1) ENOVIAVPM 管理人员 (ENOVIAVPM Manager)

ENOVIAVPM 管理人员产品 (ENOVIAVPM Manager) 为部分负责 (系统/子系统) 或全部负责计划/产品的管理人员提供一个管理环境, 以保证整个项目/计划的成功, 使用 ENOVIAVPM 管理人员产品,

企业的管理人员可以保证最终产品符合合同要求,并能按时完成,而且在工程预算范围内(标准费用和实际费用管理)。这些管理人员的工作包括:项目管理,材料清单(BOM)管理维护和产品配置规划。

(2) ENOVIAVPM 工程技术人员 (ENOVIAVPM Engineer)

ENOVIAVPM 工程技术人员产品 (ENOVIAVPM Engineer) 所提供的一系列功能组合,可非常方便地用于组织管理产品或项目组织机构部门、研究开发部门或产品,设计部门中的开发人员。这些人员部分(系统/子系统)负责或全职负责产品定义和设计。ENOVIAVPM 工程技术人员解决方案提供的功能有:定义、浏览和修改产品结构;生成或更新产品数据的属性;检人和检出产品的技术文档以及其他的功能。

(3) ENOVIAVPM 行政人员 (ENOVIAVPM Casual User)

ENOVIAVPM 行政人员产品 (ENOVIAVPM Casual User) 为需要查看产品信息、但不对其进行实际操作的人员提供有应用功能和数据。这些人员的工作包括漫游整个产品结构、编辑设计数据、发布产品数据或在合作伙伴之间进行产品数据交换。

(4) ENOVIAVPM 技术文档出版 (ENOVIAVPM Technical publications)

ENOVIAVPM 生成式技术文档出版 (ENOVIAVPM Generative Technical Publications) 产品为技术文档编写人员、管理人员和工程人员提供有可在整个产品生命周期内控制与产品相关的技术文档的功能,这些技术文档包括产品技术说明书、使用和维护手册等。该产品提供了一个环境,可快速地生成技术出版物及技术出版物的更改控制,以便保证在整个开发过程中技术出版物版本的一致性。技术文档中的元素可直接连接到存贮在 ENOVIAVPM 数据库中的零件,保证技术文档准确反应产品的更改。

(5) ENOVIAVPM 系统管理员 (ENOVIAVPM Administrator)

ENOVIAVPM 系统管理员 (ENOVIAVPM Administrator) 产品提供有 ENOVIAVPM 解决方案的全部管理功能。该产品主要针对系统管理员和负责管理 ENOVIAVPM 数据库、用户注册名、角色定义和系统调整的系统人员。这些人员的工作包括:定义数据类型、为项目分配用户以及工作机构的管理和调整。系统可根据用户的需要,对 ENOVIAVPM 提供的基本组件进行调整。通过这种方式,系统管理员就可以限制用户使用某些功能,或制定多个解决方案中功能组合的规则。

(6) ENOVIAVPM 指导者 (ENOVIAVPM Direct-

or)

ENOVIAVPM 指导者包含产品文档指导者的功能,它能够实现人工发放过程的自动化。指导者功能可允许许多人或多个部门同时审查提交的产品更改,进行电子审批或否定这些更改。电子通告功能降低了整个处理过程的时间耽搁,把工程更改所需的时间从几个星期减少到几个小时。另外,该产品还具有管理产品整个生命周期中产品配置的功能,包括维护产品定义数据、产品结构、特性和变异、及其产品的有效性。

(7) ENOVIAVPM 与 SAP 和 Baan 的连接 (ENOVIAVPM Connection for SAP and Baan)

ENOVIAVPM 与 SAP R/3 ERP/MRP 和 Baan IV 系统之间的实时、双向集成,可更紧密地组织设计环境和下一流程工程应用的联系。设计师在设计阶段就可以获得产品生产定义信息。这此信息的访问不需要退出 ENOVIAVPM 或 SAP R/3 和 Baan IV 的工作环境。(如有需求,可与 IBM 进一步联系) (Contact IBM)

(8) ENOVIAVPM CDM-PM 集成 (ENOVIAVPM CDM-PM Intergration)

该产品在 CATIA 数据管理 (CATIA Data Management) 产品和 ENOVIAVPM 中具有相同的时户界面调用其功能。CATIA 数据管理和 ENOVIAVPM 之间的同步更改,可使用户查询两者的数据库,进行产品的配置。产品部件采用图形化的树状产品结构显示,可进行整个设计的漫游。CDM-PM 集成产品允许设计师在标准的多模型工作单元中修改产品结构,管理产品技术文档、产品配置和工程更改。

(9) ENOVIAVPM 高级客户化 (ENOVIAVPM Advanced Customization)

高级客户化环境产品是一个单选的应用产品,该产品提供了一个目前最先进的完整的可视化工作环境,可把 ENOVIAVPM 指导者产品和产品文档指导者产品集成到客户化环境中。该产品具有图形化的分类多层浏览功能和完整的面向对象工作环境,可用于扩充产品数据管理功能、增加配套的功能、或集成其他的应用。直观的 GUI 可方便地编 PDM 系统的版面,以满足最终用户的需要,或生成新的版面和增加新的 PDM 功能。

(10) ENOVIAVPM 产品文档指导者 (ENOVIAVPM Product Documentation Director)

产品文档指导者产品通过所提供的可靠地访问和控制产品的数据功能,可允许整个企业在可控的方式下共享信息。包括产品设计信息、装配结构和材料清单等内容的产品信息可以通过数据仓库的方式进行存贮和访问。允许多个部门同时访问产品信息,可进

行同步工程,这样,就可减少工程周期。产品文档指导者产品所提供的产品信息存取功能,大大提高了工作的效率,降低了整个产品和工程生产的费用。

(11) CATIA数据管理 (CATIA Data Management)

CATIA 数据管理产品 (CATIA Data Management)是 VM/ESA 和 MVS/ESA 环境下 CATIA 系统数据管理的基础产品。该产品允许企业自行开发针对数据管理的客户化应用,包括工程发放和工程更改、生成材料清单 (BOM) 等非 CAD/CAM 应用。这些基于 CATIA 数据管理的客户化应用可用于存贮、管理、保护和访问产品数据。

## 5 Inte PDM 系统

IntePDM 是武汉华中软件公司 Inte 系列 CAD 软件中的一员,是一个企业级的产品数据管理系统。IntePDM 采用面向对象技术和客户机/服务器 (C/S) 体系结构,面向管理与产品相关的所有数据和所有过程的目标,通过任务驱动的过程管理,以项目管理的方式将产品数据、设计活动、人员组织和应用工具统一组织和管理起来。IntePDM 已在 863/CIMS 应用示范企业和多家工厂应用实施,是 863/CIMS 产品数据管理系统发展项目招标的唯一中标系统。

### 5.1 产品功能

#### (1) 文档管理

产品数据表现为各种形式的电子化文档,亦即计算机文件,文档是产品数据的载体,是 IntePDM 管理的数据对象。文档管理是 IntePDM 的基本管理功能,建立以数据库为基础的统一电子仓库,安全、高效地进行文档的存储、创建、变更、访问权限控制;以多种方式对文档进行检索和浏览,提供对工程图样的圈阅和批注工具;通过动态编码管理、创建文档的编码,并按编码进行文档的分类。IntePDM 的文档管理提供了一种产品数据的共享机制,使人们可以迅速得到所需要的正确数据,从而有效地减少检索信息所花费的时间,保证数据的准确性、一致性和安全性。

#### (2) 产品结构管理

通过产品结构来访问产品数据,这是另一种访问数据的方法。IntePDM 的产品结构管理对任何一个选定的项目或产品建立有关的文档或装配件之间的关系,以直观的图示方式展开整个项目、产品或选定的装配件的树状层次结构,并形成相应的物料清单 (BOM),对 BOM 表进行分类、汇总等不同的统计。

#### (3) 过程管理

IntePDM 的文档管理和产品结构管理提供了对

产品数据进行组织和编排的方法,使得产品数据易于访问和利用。在此基础上,IntePDM 的过程管理着眼于人们产生和修改产品数据的方法,管理人与人之间数据的流动,解决任务对数据的影响问题。

IntePDM 的过程管理包括任务管理和文档 workflow 管理。任务管理跟踪在项目进展过程中发生的所有事件和活动,通过定义任务确定任务的目标、内容和完成任务的人员或团队,并以直观的图示化方式进行任务的分解,确定任务的流程、任务之间的关系以及任务的启动条件,监控任务的状态,从而控制整个项目的进展。文档 workflow 管理则对特定的文档所经历的处理步骤进行管理,可以对不同类型的文档建立处理流程的模板,在设计过程中跟踪和控制产品数据变化的历程。

#### (4) 人员的组织管理

人是产品设计与开发的主体,计算机只是辅助设计的工具,人是设计过程管理中必不可少的一个要素,因此 IntePDM 既进行文档、产品结构的管理和设计过程的管理,也进行人员的组织管理。人员组织管理有静态的用户管理和行政机构管理,按企业的管理要求管理 IntePDM 所有用户以及相关的专业职能部门;也有动态的团队管理,即随着设计过程中任务的产生和分解,管理每个任务临时性的团队或执行者。

进行人员的组织管理的一个主要作用在于对产品数据即文档处理的权限控制。IntePDM 除了对人员的角色和帐号通过用户管理功能进行管理外,还对用户、机构、文档以及项目进行级别的定义和控制,并且对临时性的权限要求进行授权的管理,既严格全面又灵活方便。

#### (5) 系统管理功能

IntePDM 的系统管理功能为文档管理、产品结构管理、过程管理以及人员的组织管理提供基本的软件结构与框架,针对网络、数据库环境资源进行统一的信息存储和索引等控制,并为不同的用户提供统一的操作界面。主要包括:通信服务、数据存储、应用集成和光盘管理等。通信服务提供用户间彼此通讯的简单电子邮件和设计过程中进程之间的自动消息发送;数据存储提供数据库的查询、处理以及文件管理,使用户和应用系统与数据的物理存储相分离;IntePDM 采用一定的协议和标准集成工具软件 (如 OLE) 对各种应用系统进行封装,使处理不同产品数据的应用系统与 IntePDM 成为一个统一的整体,支持各种应用如 CAD、CAPP、CAM、文字处理、表格处理等等的数据交换和共享;IntePDM 利用光盘进行海量信息的管理,方便地对归档数据进行光盘存储管理和备份管理,并自动响应用户对数据的查询。

#### (6) 局部定制功能

IntePDM 是一个管理软件系统,它提供了一套管理产品数据的机制和实现产品数据管理的框架,而这些管理功能的实现则必须适应不同企业的特点和要求,因此提供了很强的局部定制功能,可以对特定类型的任务、文档、文档属性、编码规范、零件族分类规则、BOM 表、图样标题等进行定制,这样,不用改变 IntePDM 就可基本达到用户化,使 IntePDM 成为特定企业的本地化 PDM 系统。

#### (7) IntePDM 与 CAD 系统的集成

作为一个独立于应用的 PDM 系统,IntePDM 擅长于对产品数据的存储和查询,其主要作用在于控制数据的储存与检索、按照用户定义的属性查找数据、协调不同类型数据之间的关系、管理文件的进出、提供产品结构和配置的管理、跟踪产品开发的过程、转换不同应用之间的数据流等等,而大量的产品数据即文档则是由 CAD 及其他应用系统生成或编辑的,IntePDM 通过与 CAD 及其他应用系统的集成,使产品开发中的各种应用得到协调控制,成为一个统一的整体,从而提高工作效率,改善各小组之间的交流与合作。IntePDM 采用封装、接口等方法与 CAD 及其他应用系统的集成。封装提供了从一种应用转到另一种应用的功能,IntePDM 对可用命令行驱动的应用以封装方式集成,做到即插即用。IntePDM 能够识别应用文件,并能对应用系统发出指令,但 IntePDM 并不管理文件内部的数据。而接口是比封装更加紧密的集成,自动化程度更高。IntePDM 通过接口实现了与 InteCAD 的完全集成。IntePDM 不仅管理由 InteCAD 生成的工程图图形文件,而且实现与工程图相关的管理信息的全自动双向交换,如图样的标题栏和装配图的明细表等。图样的标题栏包含了零部件的名称、编码、材料等信息,是 IntePDM 管理的图形文档的属性,可以在 IntePDM 中创建文档时产生,并在调用 InteCAD 编辑图形时自动显示在图样的标题栏中;也可以在 InteCAD 中填写或修改,当通过 IntePDM 提交归档时,这些属性信息也一并自动录入。装配图的明细表包括零部件号、名称、材料、规格等等,反映产品的结构和构成情况。这些信息可以在 InteCAD 中生成和修改,也可以在 IntePDM 中通过文档明细表管理和产品结构管理来创建和维护,在 InteCAD 和 IntePDM 中对明细表和产品结构信息的处理是一致和统一的,即图样上明细表的更改将自动反映到数据库中,而 IntePDM 在数据库中对明细表和产品结构的更改也会自动显示到 InteCAD 的图样明细表上。这样,不需用户参与,IntePDM 和 InteCAD 就能自动交换数据及文档的管理信息,使得“同样的数据只需输入一次”,不仅提高了工作效率,而且保证了产品数据的一致性。

IntePDM 与 CAD 及其他应用系统的集成有利于产品数据在不同应用领域的重复使用,增进各种团体之间的交流与合作。这种集成通过消除冗余和易出差错的手工作业,将提高工程设计的效率,并使企业能够着手采取一些新的改革计划,如实施并行工程和协同产品开发等。更为主要的是,用户有了一种能够可靠地采集和保护许多种不同类型的、构成产品定义的数据的方法,这将是企业一笔最有价值的财富。

## 5.2 IntePDM 的特点

在技术性能上,IntePDM 具有以下特点:

1) 基于商用数据库,同时又不依赖于任何特定的数据库,实现对数据库管理系统的独立性。

2) 独立于应用软件,但能很方便地集成各种应用系统和工具软件的运行。

3) 文档信息存储在数据库和光盘中,既获得了与 DBMS 等同的安全可靠性,又在存储时间与存储容量之间达到最佳平衡。

4) 提供对工程图的标题栏和明细表的处理、版本维护、圈阅、浏览等功能,以及基于编码、关键字、文档时间等的查询功能。系统能够根据图形文档的明细表自动建立各文档之间的联系,从而将产品结构树的全貌直观地显示出来。

5) 实现以数据为核心、任务驱动的过程管理。任务流管理和文档 workflow 管理相结合,既控制项目的进展,又将任务细化到可执行状态,并保证文档必须经过确定的流程才能归档,因此能有效控制实际工作中的各种活动,尤其是有计划地进行项目实施。

6) 良好的权限管理和安全机制。除了设置不同的用户角色之外,系统对用户、文档、组织以及项目设置相应的等级,通过不同文档与用户的级别来控制权限,此外,系统的日志功能记录了对文档的所有操作,进一步保证了信息的安全性。

## 6 “金银花”系列 BEE-PDM

“金银花”系统是北京航空航天大学 720 研究所研制的基于 STEP 的 CAD/CAM 集成制造系统。该系统以 STEP 标准为基础,实现了 CAD、CAPP、NCP 等设计制造子系统的集成。“蜜蜂”产品数据管理系统(BEE-PDM)作为“金银花”系统的集成平台,管理各子系统产生的产品数据,并向各子系统提供有效的手段,实现子系统间双向的数据存取,从而为各子系统实现信息集成提供有力的支持。BEE-PDM 系统是以 PDM 理论为指导,在分析了一些著名的 PDM 软件的基础上,以“金银花”系统的产品数据集成需求为依据设计的。BEE-PDM 系统在“金银花”系统中运行良好,

在系统集成中起到了重要作用。

BEE-PDM 系统是以 PDM 理论及 CAD/CAM 系统集成的实际需要为基础设计的。该系统的设计与实现既要满足“金银花”系统产品数据集成的需要,又要具有一定的先进性和开放性,为系统进一步发展奠定基础。BEE-PDM 通过系统界面向用户提供项目管理、电子仓库、集成产品模型管理、产品数据交换、电子协作等功能。

## 6.1 BEE-PDM 功能

“金银花”系统中各个 CAD/CAM 子系统都是基于 STEP 标准的,因此为了支持这些子系统对 STEP 产品数据的访问, BEE-PDM 在系统数据存储层提供了 STEP 的标准数据存取接口 (Standard Data Access Interface, SDAI)。SDAI 是 STEP 标准的一部分,它提供应用编程接口 (API),规定了应用程序对 STEP 数据及模型的存取操作,是独立于底层存储形式的、统一的、标准的接口。通过 SDAI,各子系统可以存取和操作 STEP 数据。

项目管理是 BEE-PDM 系统的重要组成部分,是对项目进行全面的监控与管理,包括项目的任务分解,过程监控与优化,项目人员、组织、资源的管理。

电子仓库是 BEE-PDM 系统的核心,提供产品数据的检索和存储访问。用户可以向系统提交产品数据,也可以下载由系统管理的其他子系统的产品数据。该模块还提供产品数据的分类、快速检索等功能。

集成产品模型是 CAD/CAM 集成制造系统的灵魂。集成产品模型管理包括集成产品模型的存储访问,应用层各个 CAD/CAM 子系统的产品模型与集成产品模型的映射关系的建立,以及产品数据与模型的一致性检查等功能。

产品数据交换通过 STEP 中性文件实现 STEP 数据与非 STEP 数据之间的交换,以支持不同 CAD/CAM 系统(如 UG、Pro/E、北航“金银花”等)之间的数据交换。该功能是一个 CAD/CAM 系统与外界联系的重要途径。

产品配置管理是以电子仓库为底层支持,以材料清单 (BOM) 为组织核心,把定义最终产品的所有工程数据和文档联系起来,对产品对象及其相互之间的联系进行维护和管理。 workflow 管理主要实现产品设计与修改过程的跟踪与控制。因为 BEE-PDM 的使用环境是面向工作组的,因此系统中提供了电子邮件与实时通信功能,使工作组成员能够以不同的方式进行交流。例如,当 CAD 完成某个零件的设计时,可以向其它的 CAD/CAM 系统发送邮件,告知其工作情况,其

它相关系统便可以启动自己的工作。实时通信提供了白板功能,使工作组成员可以在自己的计算机前共同讨论一些问题。另外, BEE-PDM 由系统管理模块实现用户管理、系统数据的定期备份及文件服务器配置等功能。

BEE-PDM 中的产品数据都是以项目为单位进行组织的。一个项目是围绕一个产品的开发而创建的,在 BEE-PDM 中进行任何操作时,必须首先打开或创建一个项目。一个项目中管理的的信息主要有四类:组织、资源、产品与过程。组织是指参加该项目的人员组织信息,包括有哪些部门,每个部门有哪些人员,以及人员的具体情况;资源是指开展该项目所使用的资源,包括软件、硬件或其它资源;产品是指产品的一些基本信息以及产品的树状结构。项目管理具有建立与维护上述信息的功能,过程管理指建立该项目的过程模型,并对其过程进行监控。

电子仓库中的产品数据是以文件树的形式组织的。每个项目都对应一个文件树,该文件树是以项目管理模块中建立的产品结构树为基础创建的。产品中的组件或零件的不同版本的数据,都可以通过这个文件树进行有效存储。

电子仓库提供数据下载、数据提交、版本管理及快速查询等功能。用户可以通过产品结构树浏览与该产品相关的数据,并在访问权限允许的情况下,把某一数据下载到本地,也可以通过快速查询功能直接定位到所需的文件上。

当某个子系统的用户要把数据提交到电子仓库进行保存与共享时,可以通过产品结构树将数据定位到某个零部件的“CAD 数据”、“CAPP 数据”或“NCP 数据”,以创建新版本或覆盖某个旧版本。用户在提交数据时要指明数据的访问权限。

版本管理是为了有效支持工程应用而设计的。电子仓库中可以保存同一种零件的多个版本,便于用户在设计中的一些试探与回退。

## 6.2 BEE-PDM 系统特点

BEE-PDM 系统特点有:

1) 以项目为单位组织数据,并参照了 STEP 标准建立项目、产品等的信息模型。

2) 电子仓库中的各种数据以产品结构为依据,并按各个 CAD/CAM 子系统来组织存放,方便了 CAD/CAM 系统中数据的管理与定位查找。

3) 提供了 STEP/SDAI 标准数据存取接口,实现 STEP 数据的存取,并为系统进一步提高数据管理粒度,实现 CAD/CAM 系统高品质的集成奠定基础。

4) 提供了集成产品模型管理功能,该功能为

CAD/CAM 系统集成提供了有力的支持。

5) 提供了产品数据交换功能,使 CAD/CAM 系统具有很好的开放性。BEE-PDM 系统很好地支持了基于 STEP 的 CAD/CAM 系统的集成,满足了“金银花”系统的需要。

## 7 CV/Optegra

CV/Optegra 是美国 CV (Computervision Corporation) 公司的面向对象的企业级数据管理软件,是对所有的 CAD/CAM 工程应用提供完整的数据管理、保密和项目组织支持的数据管理工具。利用 CV/Optegra,可以对设计数据进行集中控制并分配和存储数据,能够确保工程师在最新版本的数据上工作,实现所有设计、制造及其他相关部门之间的信息共享。CV/Optegra 提供了不低于原来系统的正确性的前提下更快的通信和控制能力。

CV/Optegra 的系统组成:

1) Optegra Vault 模块是 Optegra 系统的核心,它有效地存储和保护产品设计中有关的全部信息。

2) Optegra Locator 模块提供了一个用户界面,用以完成日常管理工作,包括图纸定位查询、审批和发放,生成报告文件和电子文档。

3) Optegra Interface 模块提供 Optegra Vault 服务器与 CAD/CAM 软件之间的通信接口,它能预先定义一系列规则,用以识别构成一个装配或零件的数据集,以便把这些数据作为单一的单元来进行处理。

4) Optegra Navigator 模块提供一个图形化的软件工具,使工程师和项目组织者能够生成、编辑、浏览产品结构和装配关系,用层次化的树结构标述零件和子装配之间的逻辑关系。以此,用户能图形化地查询和标识零件和装配的属性,生成有关产品属性信息的详细报告。

5) Optegra Vault Customizer — API 模块允许有经验的用户将自己开发的软件与 Optegra Vault 集成在一起,构造自己的项目管理系统,帮助用户生成适合自身特点的软件界面,并确保基于用户自定义规则的整个数据管理系统的一致性。

6) Optegra Configuration Master 模块针对产品配置的所有相关特性提供存储和管理功能,配置信息可依据用户权限,由图形用户界面输入和修改。

7) Optegra TMD Lite 模块是专门为中小规模的 Optegra 用户完成工程数据管理任务而配置的软件包,是整个 Optegra 系列模块的子集,用户也可以将其与其他模块结合以进一步扩展 Optegra TMD-Lite 的功能。

## 8 ISS/InSync

InSync 是美国 ISS (Integrated Support Systems) 公司的 PDM 软件,ISS 公司是产品数据管理系统和配置管理系统软件开发商。

该系统组成有:

(1) InSync 基本系统模块 (InSync Base System)

该模块可以作为完整的产品数据管理系统来使用,也可以作为其他 PDM 系统的一部分,即作为产品配置管理系统单独使用。InSync 基本系统的采用客户/服务器体系,支持 Windows NT 和多种 UNIX 操作系统作为服务器,PC 和 UNIX 作为客户,支持 Oracle 作为底层数据库。InSync 基本系统具有很强的产品结构功能以及工作流和更改管理功能。利用集成工具箱 (ITK) 可以与其他应用软件进行数据交流。

(2) InSync 配置管理模块 (Production/Support CM)

该模块扩展了 InSync 基本系统模块的功能,InSync 的材料清单 (BOM) 对统一的产品数据具有 As-Planned、As-Built 和 As-Maintained 等三种不同的视图。

(3) InSync 集成工具箱 (ITK; Integration Tool Kit)

利用集成工具箱,用户就能在 InSync 中把其他应用程序封装进来,以便于方便地共享数据,避免数据的重复处理。ITK 在保证系统和数据安全性的前提下,使得数据能由各种应用程序调用。

(4) STEP 数据交换模块 (STEP Data Exchange)

ISS 公司作为 PDES (Product Data Exchange using STEP) 组织成员之一,STEP 数据交换模块是 ISS 公司提供的遵循 AP203 协议的、交换 AP203/CC1 数据的软件包。用户可以利用此模块,在 InSync 和其他 CM、PDM、CAD 系统之间交换各种产品数据。

(5) 用户化工具箱 (Tailoring Kit)

InSync 是通过 Tailoring Kit 来进行系统用户化的,企业可以根据自身的需要修改界面,增加新的数据单元、数据表格等,以便更好地利用 InSync 的 CM/PDM 功能。

## 9 SherpaWORK

美国 Sherpa 公司的 PDM 产品是由一系列软件模块 SherpaWORK 组成,用户可以根据各自需要来进行组合。SherpaWORK 软件模块使用了 Web 技术,利用 Internet 和 Intranet 使得整个企业范围内的所有用户都能在统一的 Sherpa PDM 环境中存取、查看、批注文

档、图样、CAD 模型等数据。

该软件的系统组成为：

(1) Sherpa/IPD (Integrated Product Development)

集成化的产品开发环境 IPD 也是实施并行工程的环境,IPD 管理产品生命周期内的所有数据,使得各小组成员能查看当前正在设计或制造的产品结构。

(2) Product Teams

这个应用模块适用于工程经理或总监,他们的主要任务是查看和审批各种产品数据。Product Teams 能够搜索、查看产品结构,查看、审批并提供各种 PDM 对象的历史信息,而且提供查看/批注功能。

(3) Sherpa/WEB

这是一种安全的 Internet PDM 浏览器,从而能够实施真正的企业级的 PDM 技术。Sherpa/WEB 是 Netscape 商业服务器和 Sherpa PDM 服务器的集成,包括 HTML 窗口和 JAVA 应用程序,在各个用户端可以设置 HTML 查询和记录窗口的配置。

(4) Sherpa/Explorer

该模块利用 Internet & Intranet 技术,提供了一个方便、可靠的产品数据存取界面。Sherpa/Explorer 允许在任何地点,在不使用原先的创建工具的情况下,存取、查看产品数据信息。

(5) Sherpa/PDM@Nct

该模块使得企业通过专用的数据线或 Internet & Intranet 来连接各使用点,用户需要用 HTML 和 JAVA 来进行配置。

(6) Sherpa/SFC (Sherpa Foundation Classes)

这是符合工业标准的对象传输协议 (CORBA 和 OLE) 的面向对象的分类库,同时也是 Sherpa/WORKS 的基础。这种技术使得 Sherpa 用户可以方便地集成其他软件,如 CAD、Microsoft Excel 等。

(7) Sherpa/WORKS 集成

Sherpa/WORKS 已经建立了针对 SFC 对象的 PDM 系统与第三方软件工具的集成“桥梁”。目前已经完成 Sherpa 和 Mentor 的集成, Sherpa 和 Pro/ENGINEER 的集成, Sherpa 和 Rosetta 的集成, Sherpa 和各种 Report Writer 的集成。

## 第4章 计算机辅助工程分析(CAE)软件

### 1 ANSYS 软件

ANSYS 有限公司由 John Swanson 博士创建于 1970 年, 该公司开发利用计算机技术进行工程分析的软件。ANSYS 有限元程序是 ANSYS 有限公司的主要产品。它开发初期是为了用于电力工业, 现在已能满足从汽车、电子到宇航、化学等大多数工业领域有限元分析 (FEA) 的需要。ANSYS 软件作为一个大型、通用的有限元程序, 其功能已为全世界所公认。ANSYS 是迄今为止世界范围内唯一通过 ISO9001 质量认证的分析设计类软件, 是美国机械工程协会 (ASME)、美国核安全局 (NQA) 及近 20 种专业技术协会认证的标准分析软件。ANSYS 是第一个通过中国压力容器标准化技术委员会认证并在 17 个部委推广使用的分析软件。

#### 1.1 ANSYS 的主要技术特点

ANSYS 软件的主要技术特点: 软件把前后处理与分析求解结合起来; 能实现多场及多场耦合功能, 实现前后处理、分析求解及多场分析统一数据库, 是具有流场优化功能的 CFD 软件; 从微机、工作站、大型机直至巨型机所有硬件平台上全部数据文件兼容; 智能网格划分; 支持从 PC、WS 到巨型机的所有硬件平台; 可与大多数的 CAD 软件集成并有接口; 有良好的用

户开发环境。

#### 1.2 功能简介

ANSYS 软件是融结构、热、流体、电磁、声学于一体的大型通用有限元分析软件, 可广泛用于核工业、铁道、石油化工、航空航天、机械制造、能源、汽车交通、国防军工、电子、土木工程、造船、生物医学、轻工、地矿、水利、家用电器等一般工业及科学研究。其主要功能包括:

##### (1) ANSYS 界面 (GUI)

基于 Motif 的图形用户界面, 各种硬件平台上界面统一, 前后处理、求解过程界面统一; 智能化菜单引导、帮助系统; 下拉式菜单; 对话框; 工具杆; 命令输入窗; 多窗口图形显示; 图形显示; 动画显示 (切片动画、等值面动画等)。

##### (2) 前处理

ANSYS 提供了实体建模及网格划分工具, 可以方便地构造数学模型及有限元模型。有多达数十种丰富的体素库可以模拟任意复杂的几何形状; 可直接建模与实体建模相结合, 自顶向下与自底向上建模相结合; 布尔运算功能; 丰富的网格划分工具可确保单元形态及求解精度; 自由网格划分、映射网格划分、智能网格划分、自适应网格划分; P 单元。



图例-4-1 某客车动力分析

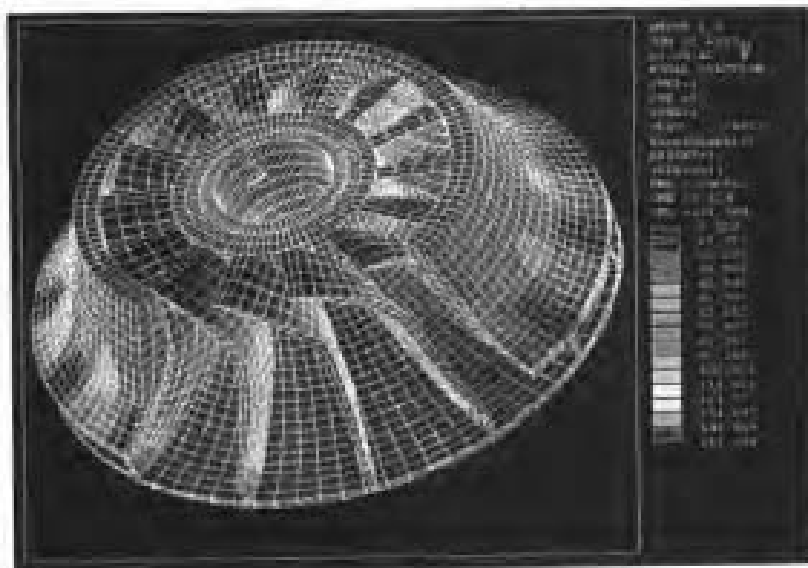


### (3) 分析功能

分析功能有结构分析、热分析、高度非线性结构动力分析、流体动力学分析、电磁场分析、声学分析、压电分析和多场耦合分析。图附-4-1为某客车动力分析。

结构分析包括：线性、非线性结构静力分析；结构动

力分析，包括模态分析（包括模态循环对称、预应力模态等）、谐波响应分析、瞬态动力响应分析、谱分析、随机振动分析等；隐式、显式（ANSYS/LS-DYNA）及显式—隐式—显式耦合求解；线性及非线性屈曲分析；非线性。图附-4-2为压缩机整体结构分析。



图附-4-2 压缩机整体结构分析



图附-4-3 电容器瞬态温度场分析

热分析包括：稳态温度场分析、瞬态温度场分析、相变分析、辐射分析、热—结构耦合分析及电—磁—热—结构耦合分析。图附-4-3为电容器瞬态温度场分析。

非线性结构动力分析包括：全自动接触分析；网格重划分（Re-meshing）；广泛的非线性材料模式（78种）；整车碰撞分析包括安全气囊分析及乘员响应分析；金属成型分析，包括滚压、挤压、锻造、挤拉、仿

似挤压、多阶段工艺、金属切割、锻压、压平、箱型成型、板壳冲压、剖面、滚压、深冲深拉等各种成型方式的模拟及回弹量计算；焊接模拟分析；高爆炸药、爆炸机理、起爆效果及应力波的传播分析；鸟撞及叶片包容分析；穿透、穿甲问题仿真；听罐及容器设计、运输、工具、运动器材及防护帽的设计；多体动力学分析。

流体动力学分析包括：层流分析；湍流分析；自由

对流与强迫对流分析；可压缩流/不可压缩流分析；亚音速、跨音速、超音速流动分析；多组份流动分析（达6组份）；移动壁面及自由界面分析；牛顿流与非牛顿流体分析；内流和外流分析；化学反应；分布阻尼和FAN模型；热辐射边界条件；管流。

### 1.3 ANSYS7.0 新增功能

整体新增功能——ANSYS Workbench Environment

1) 为一可选配的使用者接口，增强在进行简谐、大变形与多负载组合分析的控制能力。

2) 在单一装配组立件上，能同时进行实体与曲面模型的网格化动作。

3) 可将分析结果输出至 Microsoft Excel。

4) 分析得到的反作用力的资料可显示在详细菜单中，并可设为参数。

5) 支持 SpaceBall 与 SpaceMouse 的 3D 运动控制。

6) 工作表浏览器可以像 Excel 那样，以文书夹的形式方便地显示对象的摘要内容。

7) 可依照使用者指定的方式，显示直角坐标或是圆柱坐标系统的结果。

#### (1) 求解器新增功能

1) Sparse 求解器可以在次结构分析的产生过程中使用，在扩展过程中完成之前的代换。

2) 复合非对称 Sparse 求解器能使用在完全简谐分析的平行处理计算上。

3) 经由 PCG 求解器的内存节省选项，SOLID95 (SOLID186) 元素能与 SOLID92 (SOLID187) 元素一起使用。

4) 分布式领域求解器 (DDS) 支持拘束方程式与耦合计算，且可以使用在单一接触分析上。

#### (2) 加强易用性

1) CMLIST、CMPLOT 和 CMSEL 指令可减少组件的使用名称与使用的说明。

2) 新的 NRRES、ESCHECK 和 CNCHECK 指令可在非线性结构分析中，与修订过的 CHECK。

#### (3) 结构分析新增功能——新增元素

1) 在工程应用方面，MPC184 元素在两个变形体或刚体之间提供刚体拘束，来传递力或力矩。

2) SHELL131 (4 节点) 和 SHELL13 (8 节点)，3-D 层元素，在平面内或厚度方向具有热传导功能，并提供热梯度结果。

3) CONTA175，一个点/边对面对面的接触元素，可在节点与曲面间，或是边线间模拟接触与滑动。

#### (4) 元素新增功能

1) 复数形式的压力负载，可以施加在 SURF153 和 SURF154 元素。

2) LINK180、BEAM188 和 BEAM189 可支持黏弹性的分析。

3) 对于 PLANE182 和 PLANE183 元素，可使用平面应变选项。

#### (5) 新增材质模型

超弹性材质模型 Gent 与 Yeoh，和 Blatz-Ko 与 Ogden (泡沫材料) 一样，用于模拟高度压缩性的弹性体。

#### (6) 指令新增功能

1) 利用 CMOMEGA 和 CMDOMEGA 指令，定义旋转元素组件的角速度和角加速度。

2) 对于个别的超弹性材质模型，新的实验曲线的拟合能力可用来比较经由实验得到的应力数据，与计算分析出来的结果。

3) 经由 KEYOPT 的设定或是接触精灵导引，对于曲面对曲面接触元素，可仿真导电接触现象。

#### (7) 热传分析新增功能

1) 改善辐射问题的求解时间。

2) 辐射视界系数 (View Factor) 可以二进制格式储存。

#### (8) 耦合场分析新增功能

1) 降阶建模 (ROM) 工具可求解包含挠性结构的耦合场问题

2) 2D CIF 格式档案的输入，可用于求解模拟多层集成电路组件

#### (9) 电磁分析新增功能

1) 对于 S-参数计算，提供新的指令或宏；

TLSPRM 宏指令；用来计算一般已知输入阻抗的 n-port 网络的 S-参数；

SPSWP 宏指令，用来计算在一个指定频率范围内的 S-参数；

PLSP 指令，用来将 S-参数在 XY 平面上绘制曲线。

2) HFEREFINE 宏指令用来执行 HF119 元素的自适应网格。

3) SOLID91 和 SOLID117 元素新的主要选项，确定电流密度的螺旋线圈环境。

#### (10) 流体分析新增功能

1) 新的温度场算法，用来计算共轭热传问题的表面薄膜系数。

2) CFD 分析可输入局部表面粗糙度。

#### (11) LS-DYNA 新增功能

1) 新的 EDAL 和 EDSCALE 指令，支持对于特大变形问题(如挤出或锻造)所采用的任意拉格朗日欧

拉 (ALE) 算法。

2) 在 HP Alpha 服务器平台上, 新的 EDDBL 指令能够使双精度功能用于连续的显性到隐性回弹分析。

3) ANMRES 指令, 允许在多个结果文件间产生动画。

(12) EMAX 新增功能——模型

1) 支持 IGES import 功能。

2) 加强的对象调整工具列。

3) 新的几何原素: 波导、抛物面天线、螺旋与环状线、点。

4) 增强 CAD 模型输入, 允许曲面集合分配给任意曲面上。

5) 曲面集合可分配到所有 CAD 模型上的退化表面上。

(13) 计算

1) 新的快速扫频工具, 可在一个频段内快速萃取 S-参数。

2) 加强的天线计算功能, 包括增益、辐射功率、方向性和效率。

3) 新的 RCS 极化选项。

4) 自动计算 n-port 系统的 S-参数。

5) 支持 S-参数的标准文件。

(14) 速度与效率

可藉由减少输出计算以提高求解效率。

(15) 用户接口

1) 重新设计的 GUI 图示。

2) 增强了工具列功能。

## 2 工程分析软件 NASTRAN

### 2.1 结构分析软件

MSC/NASTRAN 是世界上功能全面、应用广泛的大型通用结构有限元分析软件, 同时也是工业标准的 FEA 源代码程序及国际合作和国际招标中工程分析和校验的热门工具, 它可以解决各类结构的强度、刚度、屈曲、模态、动力学、热力学、非线性、(噪) 声学、流体-结构耦合、气动弹性、超单元、惯性释放及结构优化等问题。

NASTRAN 程序提供开放式用户开发环境和 DMAP 语言及十余种 CAD 接口, 以满足用户的特殊需要。

MSC/DYTRAN 主要用于求解高度非线性、瞬态动力学、流体及流固耦合等问题, 可解决广泛复杂的工程问题。如: 金属成形 (冲压、挤压、旋压、锻压)、爆炸、碰撞、搁浅、冲击、穿透、汽车安全气囊 (带)、液

-固耦合、晃动、安全防护等问题。程序采用有限元方法及有限体方法, 并可二者混合使用。

MSC/FATIGUE 是专用的耐久性疲劳寿命分析软件系统, 可用于零部件的初始裂纹分析、裂纹扩展分析、应力寿命分析、焊接寿命分析、随机振动寿命分析、整体寿命预估分析、疲劳优化设计等各种分析。同时该软件还拥有与疲劳断裂有关材料库、疲劳载荷和时间历程库等, 对分析的最终结果具有可视化特点

MSC/Construct 是基于 MSC/PATRAN 和 MSC/NASTRAN 用于拓扑及形状优化的概念化设计软件系统。通过该软件可根据设计性能预测改变结构材料分布, 构造新的拓扑关系和几何特征, 并进而通过非参数形状优化光滑拓扑优化模型降低应力级别, 提高产品设计寿命。MSC/Construct 可在网格自适应技术的基础上实现网格重划分功能, 处理多种载荷及边界条件, 对解决超大型模型同样有效。

MSC/MARC 是功能齐全的高级非线性结构有限元分析系统。它具有极强的结构分析能力。可以处理各种线性和非线性结构分析包括: 线性/非线性静力分析、模态分析、简谐响应分析、频谱分析、随机振动分析、动力响应分析、自动的静/动力接触、屈曲/失稳、失效和破坏分析等。可以解决各种高度复杂的结构非线性、动力、耦合场及材料等工程问题。尤其适用于冶金、核能、橡胶等领域。

MSC/AKUSMOD 是 MSC 公司与德国 SFE 公司共同开发的内噪声预测仿真软件。该系统以 MSC/NASTRAN 为基础, 通过 3D 流体网格自动生成, 自动地完成流体结构耦合。先进的吸波单元和全面的可视化技术, 可进行振动噪声分析、内噪声预测及噪声优化灵敏度分析等, 适用于汽车、航空、铁路、船舶各个领域。

### 2.2 建模、有限元前后置处理软件 MSC/PATRAN

MSC/PATRAN 是集几何访问、有限元建模、分析及数据可视化于一体的新一代框架式软件系统。MSC/PATRAN 独有的 SGM (单一几何模型) 技术可直接在几何模型一级访问先导的 CAD 软件数据库系统包括 CADD5, UG-Euclid, Pro/ENGINEER, CATIA 等, 以及 I-DEAS。除了框架系统外, MSC/PATRAN 还拥有功能强大的各种专用分析应用模块, 主要包括:

1) MSC/PATRAN ADVANCED FEA。MSC 与世界一流非线性 CAE 软件供应商 HKS 共同开发的高度非线性分析软件, 其包含了 HKS/ABAQUS 的大部

分功能,可解决各种静力、动力、热传导、接触、非线性等高度非线性问题。

2) MSC/PATRAN FEA。提供了静力、动力、热传导、非线性及优化灵敏度分析等功能,可满足各类工程需要的通用结构有限元分析软件。

3) MSC/PATRAN THERMAL。专用工程自动热分析软件系统,可处理各种复杂的热学问题,并提供了丰富的与温度相关的材料库,对流关系库及与时间和温度相关的专用函数库,包括角系数计算。

4) MSC/PATRAN LAMINATE MODELER。专用的集复合材料设计,分析,制造于一身的软件系统,可帮助工程师随意构造所需的复合材料模型,进行损伤计算或通过有限元模块完成结构分析,为生产提供相关的复合材料制造参数及材料加工数据。

5) MSC/ARIES 作为 MSC/NASTRAN 面向设计和分析,易学易用。开放的前后处理器, MSC/ARIES 提供了先进的三维实体建模,零件装配,参数化设计,有限元网格自动划分,材料特性定义,机构运动仿真,可视化处理等多种功能及集成的 CAD/CAE 环境。包括与其他各种 CAD 软件的接口。

### 2.3 集成的前后处理及工程分析软件

MSC/NASTRAN For Windows 是运行在 Windows NT 环境下 PC 机及笔记本电脑上的 MSC/NASTRAN 产品。它采用集成化的前后处理器,配合先进的高级实体建模功能,可极其方便地进行结构应力分析、模态分析、屈曲分析、稳态或瞬态热分析、非线性结构分析、动力响应分析、结构优化及 CFD 高级流体力学计算等。并直接与 AUTOCAD、AutoDesk MDT、SolidWorks 等连接运行,同时通过专门接口访问其他 CAD 系统。

### 2.4 嵌入到 CAD 软件中的 MSC 分析软件

MSC/InCheck 是专门面向设计人员的设计校验有限元分析软件系统。其采用 MSC/NASTRAN 的技术核心,可与 AutoCAD、AutoDeskMDT、SolidWorks、SolidEdge 等环境实现无缝集成。MSC/InCheck 全智能化的图形引导器无须设计人员拥有专门的有限元知识,即可实现线性静力、模态、屈曲等强大的分析功能。

## 3 机械系统动力分析软件 ADAMS

机械系统自动动力分析软件 ADAMS (Automatic Dynamic Analysis of Mechanical Systems) 是世界上最广泛使用的机械系统仿真 (Mechanical System Simulation—MSS) 软件。通过预测和分析机械系统

经受大位移运动时的性能,ADAMS 可以帮助改进各种机械系统的设计:从简单的连杆机构到车辆、飞机、卫星、洗衣机、VCR 机构,磁盘驱动器甚至复杂的人体。

ADAMS 为工程师提供各种生成并试验其设计方案的途径,这在以前是不可能做到的。ADAMS 软件样机能够在物理样机和试验数据得到前很久即可进行完整的系统仿真。其他各种可供选择的设计方案也可进行仿真试验、修改和优化;这些都可以在设计过程的早期进行,这样就大大降低了成本、极大地缩短了新产品投入市场所需的时间。

### 3.1 ADAMS 的特点

分析类型包括运动学、静力学、准静力学分析,以及完全非线性和线性动力学分析;具有 2 维和 3 维建模能力;包含刚体和柔性体分析;具有 50 多种联结副、力和运动发生器组成的库;具有组装、分析和动态显示不同的模型或同一个模型或同一个模型在某一过程变化中的能力;具有一个强大的函数库供用户自定义力和运动发生器;具有开放式结构,允许用户集成自己的子程序;具有先进的数值分析技术和强有力的求解器,使求解快速、准确;具有与 CAD、FEA、Rendering (广告动画) 和控制系统建模软件之间的专用接口;具有易使用的图形界面 ADAMS/View。

### 3.2 ADAMS 的使用范围

1) 汽车工业:用于车辆的乘坐平稳性、可操纵性和寿命研究;悬架系统和转向性能分析;驾驶训练动力学研究以及各种各样机构设计。

2) 航天工业:用于卫星结构的展开、卫星轨道及其飞行姿态动力学研究。

3) 航空和国防工业:用于飞行稳定性和控制分析,驾驶员弹射模拟分析,装备分离、起落装置和载荷分析。

4) 工程机械行业:用于车辆越野机动性分析,操纵性和寿命分析,挖掘机、起重设备以及卡车的动力学分析仿真和研究。

5) 机电产品工业:用于激光唱机、VCR 机构、照相机。杆件机构和照片复制机的分析。

6) 生物力学和人机工程领域:用于乘员碰撞仿真分析和乘员姿态分析,以及人机界面的检验和事故重建。

### 3.3 ADAMS 的建模技术

在 ADAMS/View 模块中,使用积木法 (building-block) 和操作方便的图形界面,用户可通过指定下面

的四种基本元素来建造机械系统仿真模型,即部件、约束、运动发生器和力。部件是系统中的组成零体,它具有几何形状、质量和惯性特性,结构中的柔性和阻尼是通过离散结构单元,并输入相应的单元刚度值来实现。约束是指联接系统中的各部件,并对运动给予约束限制。运动发生器(包括平动和转动)是用来控制联接副的运动,它是一个随时间变化的函数。

广义力和力矩可以下列各种形式包含在系统建模中:弹簧阻尼器;单方向力或力矩分量;多个分力或力矩分量;力或力矩矢量;梁(无质量、等截面线性单元);衬套(具有刚度和阻尼的六个分力或力矩);线性力场(任意的 $6 \times 6$ 阶刚度和阻尼矩阵)。

由于有了线性力场,控制建模单元就可加入到系统模型中,模拟系统对已定义的控制系统的响应。这些控制建模单元包括:传递函数;线性广义状态方程,用于表示线性和非线性、多输入/多输出控制规律。

### 3.4 ADAMS 软件的用户化

通过使用函数表达式和用户子程序,可以使 ADAMS 很容易适用于特殊的应用。这些特点给用户提供了一种手段来定义 ADAMS 模型中诸状态变量之间的符号关系。由函数表达式或用户子程序计算出的瞬态值可赋给:约束值和运动输入;力和力矩分量;用户定义的数值变量;辅助输出量;

传感器(位于模型中,测试感兴趣工况的动态特性)。

函数表达式和用户定义的程序可取决于下述状态变量和函数:运动学函数,如用户感兴趣的点的位移、速度或加速度;动力学函数,如作用力和反作用力;用户定义的数值变量和动态变量;FORTRAN-77 内部函数,如 SIN、ASIN 等;曲线、曲面的插值函数;控制单元的输入、状态和输出;正交函数,如多项式、富利哀级数等;常用的平滑函数、谐波函数等。

函数式直接在 ADAMS 数据集中定义,并在分析阶段求值。用广定义的程序可编译并直接连接到 ADAMS 库中以产生一个 ADAMS 专用用户版。

### 3.5 ADAMS 分析功能

ADAMS 分析求解的第一步是初始条件(Initial Conditions)和过约束(Redundant Constrain)分析,分析初始模型配置,检查该模型是否已正确组装,然后采用优化技术调整该系统的配置,确保其正确性和可行性。分析中若存在过约束,则将其消除。这些功能极为有用,因为它们可以允许用户修改部件设计而由 ADAMS 软件本身自动确定由此而产生的组合系统的变化。模型一经校核,用户就可以进行运动学、静力学、

准静力学以及动力学分析。

运动学分析就是计算系统的位移、速度和加速度。此系统所有自由度受系统配置和运动约束的支配;静力学分析就是确定系统的静平衡位置和此时的相互作用力,对于确定系统上的初始载荷,其静态解是很重要的,它提供给用户一种获得系统平衡的方法,并能在系统动力学分析时消除不希望的瞬态起动效应;准静力学分析的求解技术同静力学分析,但此时用户可以指定一系列连续配置状态进行分析;动力学分析就是全面预测多自由度系统对惯性效应响应时的运动和力。

### 3.6 ADAMS 参数化分析功能

ADAMS 允许用户交互式修改多个设计参数,如系统质量、刚度、阻尼系数和联接副位置等,分析在某一时段,其模型或一系列模型因参数的改变而响应的变化情况。

ADAMS 提供了一系列命令,使上述过程简单易行。Activate 命令和 Deactivate 命令的联合使用可建立某一模型,以满足其模型参数在一系列变化时特殊的工作或系统配置的要求,Save 命令配以 Reload 命令用来恢复任何贮存的状态以便 ADAMS 自动从同样的位置进行多状态仿真分析。File 命令使用户能够在单一的 ADAMS 执行中读取多个模型并进行分析。

## 4 结构有限元分析与优化设计软件系统 JIFEX2.0 版

JIFEX 是由中国大连理工大学工程力学系/工程力学研究所研制开发的新一代有限元分析与结构优化设计软件系统。它是在多层子结构有限元分析程序 JIGFEX、微机有限元分析软件 DDJ-W、计算机辅助结构优化设计软件 MCADS 等自主知识产权的基础上发展起来的具有新的前后处理界面的集成化软件系统。JIGFEX 的发展始于 20 世纪 70 年代,是中国最早开发出来的有自主知识产权的大型通用有限元分析软件,于 1981 年通过了由中国国家教育部组织的技术鉴定,其后投入工程应用,在运七飞机、气垫船、直线粒子加速器、重庆长江大桥等许多重要工程结构设计分析中应用。

自 1981 年推出 JIGFEX 第一个版本以来, JIGFEX 软件系统一直是中国计算力学与 CAD 研究领域最知名的具有自主知识产权的有限元分析软件之一。在其后十几年时间里, JIGFEX 软件系统在不断应用中不断发展,并相继发展了一批分支软件,如微机通用有限元分析软件系统 DDJ-W、海洋石油平台设计分析软件 DASOS-J、高层建筑结构设计分析软件 DASTAB、屈曲稳定分析软件 DDJTJQ、建筑结构计算机辅助设

计软件 FCAD、结构优化设计软件 DDDU、计算机辅助结构优化设计软件 MCADS、JIGFEX 的微机版本以及结构与上动力学及渗流分析软件 DIASS 等。这些软件通过了有关部门的技术鉴定,在土木、机械、国防、航空、航天、车辆、石油、化工等广泛的工业领域中应用,在经济建设和国防建设中发挥了重要的作用。相应的一批研究应用成果获得了国家与部委级的科技奖励。其中一些软件也成为商品化软件推广,在中国建立了比较广泛的用户群。JIGFEX 系统及其分支软件,不仅在中国享有声誉,而且得到了国际同行的认可与重视,在国外的文献与数据库中给予介绍和评价。

#### 4.1 JIGFEX2.0 的前置处理子系统 Auto FEM

JIGFEX2.0 的有限元建模前处理子系统 Auto FEM,是以工业界普遍采用的 CAD 软件 AutoCAD (R14, R13) 为平台的有限元建模软件。它利用 AutoCAD 优秀的建模功能,在 MS Windows/Windows95/Windows NT 环境下方便灵活地人机交互生成有限元计算模型,自动生成全部有限元模型数据。从而为用户提供了全新的有限元前置处理界面,极大地简化了有限元建模这一困难的工作,并将有限元分析与 CAD 产品模型设计连为一体。

JIGFEX-AutoFEM 的主要功能有:

- 1) 人机交互地构造结构与产品设计的几何模型。
- 2) 将结构与产品的几何模型离散化,选用下列方法自动生成有限元网格:自由网格法、映射法、扫描法、3D-MESH 法、编织法;可以指定边界的节点分布,局部加密网格,对节点优序编排。
- 3) 交互式生成有限元计算模型的属性数据,包括材料参数、单元截面特性、外部载荷,边界约束条件等。
- 4) 交互式建立各种截面形状及任意空间位置的框架结构模型,并自动生成梁单元的刚度特性数据。
- 5) 有限元模型的编辑与修改,包括部分区域的网格及属性数据的平移、旋转、拷贝、镜像、阵列、更新、删除。
- 6) 有限元模型的真实图形显示与查询。
- 7) 生成有限元分析模型通用数据库(Universal 文件),通过转换程序生成有限元分析软件前端输入数据文件。

#### 4.2 JIGFEX2.0 的图形后置处理和视算一体化子系统 GRAFE

(1) 全新的 Windows 界面

1) 使用标准的工具条、状态行,对所有按钮均提供提示功能,使用方便快捷。

2) 使用方便的右键快捷菜单,为用户提供当前可使用的所有操作。

3) 灵活的树形目录管理功能,分类管理各种数据,方便数据的检查与管理。

4) 交互式多窗口计算机图形显示功能,将各种图形同时展现在您的面前,各窗口内的图形独立操作。

(2) 强大的图形处理能力

多种有限元模型及其计算结果数据的图形显示:有限元模型图、结构变形图、三维模型消隐、实体光照模型、等值线图、应力彩色云图、结构振型及响应的动态图形模拟等;各种图形可在不同的窗口内显示,也可在统一窗口内切换,操作方便;实时的图形几何变换功能,可对各窗口内图形进行实时的旋转、平移、缩放等各种几何变换;方便的参数设置功能;通过在不同的控制卡片内设置相应的参数即可控制所显示的内容、显示的方式等;基于 Windows 系统的图形打印功能,只要 Windows 支持的打印机均可打印。

(3) 可扩展的图形与数据接口

1) 输出矢量和点阵图形,可嵌入字处理软件如 Word 中进行图文编辑。

2) 标准的图形数据接口。输入规定格式的图形数据;对于用户的数据文件格式,只需开发相应接口程序即可输入,实现与各种有限元软件的连接。

(4) 分布式计算可视化

1) 基于动态数据交换(DDE)的实时的计算可视化能力,计算与图形显示同步进行。

2) 基于 TCP/IP 协议的分布式计算能力。可将计算任务分布到不同的计算机中进行,提高计算与图形处理效率。

#### 4.3 JIGFEX2.0 有限元分析的结构模型化特点

(1) JIGFEX2.0 结构模型化功能

JIGFEX 系统具有各种类型单元和多种荷载模式模拟各种复杂结构,灵活的节点位移规格、任意的局部座标系和多重主从位移控制等功能适用于处理特殊的结构模型。其所带来的优点是:一方面使诸如结构中的刚性构件与刚性连接等各种结构模型得到精确的处理;另一方面使工程计算中常见的不同类型有限单元之间的连接(如梁、板;膜连接,梁、板;块体连接等)非协调一类问题得到很好的处理。

(2) JIGFEX2.0 的多层子结构模型化功能

采取多层子结构方法对结构进行模型化是 JIGFEX 的一个显著特点,该方法是将一个规模较大的复杂结构分解成一系列规模较小的子结构来处理。在计算上实际是将一个阶数十分高的方程组转化为一系列阶数

较小的方程组来进行求解,且各组方程之间具有相对的独立性。其优点是可以简化有限元建模、提高计算速度、节省计算机存储、简化局部非线性和接触问题计算工作量等。

#### 4.4 JIFEX2.0 的各有限元分析子系统

(1) JIFEX2.0 的弹性静力有限元分析子系统 FeaSTAT

弹性静力分析是工程结构分析与设计的基础。弹性静力分析所涉及的问题也是多方面的。JIFEX-FeaSTAT 具有各种有限元静力分析功能,其中包括热应力计算、轴对称结构和复合材料结构分析等等。JIFEX-FeaSTAT 具有求解超大规模复杂问题的能力,对于十万至二十万量级自由度规模的分析问题,在常规的微型计算机上就可进行求解。JIFEX-FeaSTAT 的复合材料模型包括层合板和夹层板。

(2) JIFEX2.0 的动力有限元分析子系统 FeaDYNA

JIFEX-FeaDYNA 的动力分析包括固有模态、频率响应、时程响应、地震响应分析等。JIFEX-FeaDYNA 采取迁移子空间迭代法可快速准确地求出几阶或几十阶频率。瞬态分析可以计算在随时间变化的荷载作用下结构的响应。频率响应计算,可以考虑调谐激励荷载和基础加速度运动对结构响应的影。例如,计算人员可以通过频率响应计算来预测结构最大的响应位置。在自由振动分析之后,输入少量数据即可完成对二维或三维结构(如桥梁、水坝、楼房等)的随机地震响应分析。

(3) JIFEX2.0 的屈曲稳定性有限元分析子系统 FeaBULK

适用于杆、梁、板壳、复合材料板壳以及其他各种单元的组合结构屈曲稳定性分析。JIFEX-FeaBULK 系统可以用静力有限元分析的结果作为组合结构的初始前屈曲膜力进行弹性屈曲分析,也可根据对给定的初始前屈曲纯膜力状态进行屈曲分析;其中后者根据简单结构膜力能用简单公式算出的特点,将膜力作为原始数据输入,从而节省计算量。

(4) JIFEX2.0 的热传导有限元分析子系统 FeaTHERM

JIFEX-FeaTHERM 的热传导分析包括稳态与瞬态热传导的温度场计算。许多工程应用中,解决热传导问题是非常必要的,它也是进行热应力分析的第一步。JIFEX-FeaTHERM 用单元网格模拟任意几何形状,温度数据可直接输入到有限元分析中,可选择固定时间步长和自动时间步长,当接近稳态时,时间步长将增加;反之,如果材料特性或边界条件发生剧烈变化时,

时间步长将减少。

(5) JIFEX2.0 的接触有限元分析子系统 FeaCONT

JIFEX-FeaCONT 的接触有限元分析功能包括有摩擦、三维接触、多体接触、弹塑性接触计算。其适用于机械、土木、航空等工程实际中的大量接触问题的求解。例如轴承系统中滚子与滚道、外圈与轴承箱的接触,齿轮之间的相互咬合、接触,火车轮-轨系统在滚动过程中的摩擦接触,岩土工程中土与墙、不同岩层间的弹塑性接触,结构与土基的耦合作用接触等等。JIFEX-FeaCONT 通过子结构描述将接触分析置于最高层子结构,从而可以处理大规模复杂问题,而且以参变量变分原理和参数二次规划算法为基础的非线性计算过程收敛迅速平稳,并确保了问题的计算精度。

(6) JIFEX2.0 的弹塑性有限元分析子系统 FeaPLAS

JIFEX-FeaPLAS 适用于二维平面和三维块体以及杆单元组成的结构弹塑性分析。程序装备了工程上常用的弹塑性本构模型,如: Tresca、Mises、Coulomb 模型等。同样采用了参变量变分原理和参数二次规划求解算法,收敛过程平稳。程序配有适用于不同拉压模量杆件分析的求解器,只要问题有解,程序总能得到收敛解。

#### 4.5 JIFEX2.0 的结构优化分析与设计子系统 OPTSYST

JIFEX-OPTSYST 软件系统的结构优化设计部分是在原来的计算机辅助结构优化设计软件 MCADS 基础上集成和发展而来。原 MCADS 就是一个通用性的多单元、多任务况、多变量、多约束、多目标结构优化软件,能够完成各种复杂组合结构在静动态响应下的优化设计。在 JIFEX-OPTSYST 新版本中,MCADS 子系统又进一步发展了结构布局优化和屈曲稳定性优化等功能,并同 JIFEX 软件系统更强大的有限元分析以及前后处理功能集成,为优化设计功能的进一步发展奠定了基础。

JIFEX-OPTSYST 子系统的结构优化计算与设计功能:结构的单元截面尺寸优化设计;结构的边界形状优化设计;布局优化设计;复合材料层合板与夹层板结构的铺层优化设计;结构的重量优化设计;结构的强度和刚度优化设计;结构固有振动频率优化设计;结构的振动响应优化设计;结构的屈曲稳定性优化设计。

## 5 机械系统动力学、运动学分析软件 DADS

### 5.1 DADS 概述

DADS 是著名的机械系统动力学、运动学分析软件,能对机械系统整体的机械特性进行仿真。DADS 多年来一直应用于高端领域,如航天、航空、国防、铁道、特种车辆、轮船、汽车、机器人、生物医学等,被认为是动力学和运动学仿真方面的权威软件。典型的应用包括:航天器飞行控制及交会对接、卫星天线伺服控制(太阳能帆板展开、导弹发射及飞行动力学分析、引信、激光吊仓、伺服稳定平台、飞机起落架、车辆操纵稳定性和行驶平顺性、机器人行走稳定性、人工关节的动力学分析等。

目前, DADS 与 LMS 公司的疲劳分析软件 FALANCS、优化分析软件 OPTIMUS、噪声分析软件 SYSNOISE 及数据压缩软件 TecWare 有了专用接口。

与 DADS 有专用接口的软件有: MATRIXx、Pro/E、I-DEAS、CATIA、SolidWorks、UG、AUTOCAD、MATLAB、EASY5、ANSYS、MSC/NASTRAN、ABAQUS、FALANCS、OPTIMUS、SYSNOISE、TecWare、PolyFEM、Elfini, DADS 仿真得到的结果还可以用到其他第三方软件中。

DADS 的特点有:

- 1) 运算稳定,不会出现积分发散现象,对所有正确的虚拟样机都能解算并能得到正确的结果。
- 2) 柔性体技术最成熟。最先将柔性体技术加入到动力学仿真软件中,并得到专家认可。
- 3) 能精确仿真多柔性体机构。
- 4) 与控制系统仿真软件能完美结合。最早在内核中加入与控制系统仿真软件的接口,自动实现系统仿真,数据传递快,效率高,仿真精度好。
- 5) 易学易用。建立虚拟样机的操作简单,软件实用性强,适合在各个相关领域应用。
- 6) 效率高。虽然一直沿用 Motif 界面,但软件提供了各种措施提高建模效率。
- 7) 应用领域更广。只要有运动存在,就有可能用 DADS 建立虚拟样机。

### 5.2 DADS 模块介绍

所有的 DADS 模块都是独立的,用户可以根据需要选择有关的模块进行组合。

#### 5.2.1 基本组合 DADS/Basic

DADS/Model, GUI, Animation † DADS Solver

† DADSGraph Postprocessor, 这种组合称为基本组合 DADS/Basic。

(1) DADS/Model, GUI, Animation 建立仿真模型及动画显示仿真结果

1) Analysis 菜单包含有 System、Dynamic、Inv Dyn、Kinematic、Static、Preload。System 可以定义分析类型。分析类型有:动力学分析 Dynamic;逆动力学分析 Inv Dyn;运动学分析 Kinematic;静力学分析 Static;预载荷分析 preload。另外在 System 菜单中可以定义单位制、输出文件的类型、仿真时间及积分步长、重力加速度方向、重力加速度系数、是否进行交互式仿真等功能。Dynamic 菜单可以定义进行动力学分析的参数,包括:作用力计算开关、输出作用力值的参考坐标系、输出结果的精度、积分精度、积分方法、解算加速器的类型等。Inv Dyn 菜单可以定义进行逆动力学分析的参数。逆动力学分析一般用来计算主动轴上的输入力矩或主动件上的驱动力,要输入的参数与“动力学分析”中的参数相似。Kinematic 菜单可以定义进行运动学分析的参数。要输入的参数包括:积分精度、积分方法等。Static 菜单可以定义进行静力学分析的参数。要输入的参数包括是否在计算时加入“扰动力”等,如选择“true”,则能大大提高积分的稳定性。Preload 菜单可以定义进行预载荷分析的参数。预载荷分析的主要作用是计算机构中弹性元件在初始装配位置时内部的作用力。要输入的参数包括:力的数值误差范围等。

2) 用“Body”菜单可以在 DADS 中建立三维机械模型中的每一个零件。包含的子菜单有:“create”创建、“edit”编辑、“delete”删除、“copy”拷贝、“Enable”起作用、“Disable”不起作用。

“create”创建的功能有:

- a. Parent 父级元素,决定新创建元素的位置。
- b. Method 坐标变换的方法,有: Bryant angles (XYZ)、Euler angles (ZXZ)、Eulerparameters、PQR、Slider 五种方式,在这几种方式之间切换时,坐标转换是自动完成的。
- c. EditLabel 定义模型在动画显示时,同步显示动力学数据的数值大小、单位等。
- d. Center of Gravity 显示零件的重心位置。
- e. Geometry 给定零件的几何形状,有 Create、Edit、Delete、Use 子菜单。
- f. “Use”可以用来拷贝已有的几何形状或调入以 SLT 格式存储的文件来生成零件。在“Create”中可直接调用的几何形状库有: Box 二维矩形框、Sphere 球、Cube 立方体、Circle 圆、Cone 圆锥体、Cylinder 圆柱体、Tube 圆管、Wedge 角楔、Pyramid 棱锥、Gear 齿



轮 (直齿轮、斜齿轮、伞齿轮)、Spring 弹簧、Polygon 多边形、Extrude 拉伸体、Sweep 旋转体。

g. Triads 作用点, 定义在模型中或零件上的点, 所有这些点都带有一个局部坐标。Triads 用来连接其他零件、给零件加约束或加上外部载荷等。

h. Mass Properties 用来显示或定义材料特性。

i. Body Forces 在零件上施加载荷。在零件的质心上施加作用力; 在零件的质心上施加力矩。

j. Animation Features 在动画显示计算结果时, 用变化的“箭头”来形象地显示零件的速度、加速度、角速度、角加速度、运动副及零件连接处的相互作用力; 用“箭头”的长短来表示这些量的数值大小; “箭头”的方向即就是这些量的方向。

3) 建立局部坐标系的目的是提高在 DADS 中建模的效率。这一功能如用的好, 则模型的结构层次分明, 易于理解, 在修改模型时将十分方便。

4) 在建立 DADS 模型时, 当要用到一些非特殊的点 (如零件的质心或 Triad) 时, 必须要建立辅助点。可以选取辅助点作为零件的原点、Triads 的原点或作为局部坐标系的原点。在定义了 Point 后, 则在一些 ORIGIN 后, 既可以选择 Numeric, 也可以选择定义好的辅助点作为原点。

Coordinates 输入坐标值

EditLabel (

Parent

5) Constraint 约束。定义模型中某些零件之间的关系或某个零件以一定的规律运动, 可以用菜单 Constraint 实现。

a. Diff 两个零件之间在 X、Y、Z 方向的位置关系 (各种关系都是以零件上指定的 Triad 为参照点确定的)。

b. Dist 两个零件之间的位置关系 (即两个零件上指定的 Triad 保持一定距离)。

c. Dot1 第二个零件的轴线与第一个零件的两根轴线垂直。

d. Dot2 第一个零件上的某一点与第二个零件上的某一点之间的连线与第一个零件的两根轴线垂直。

e. Driver 给定运动规律: One-Body 某一零件的绝对运动规律; Two-Body 两个零件之间的相对运动规律; Gear 齿轮传动中其中某一齿轮的运动规律; PntCurve 在点为曲线接触时运动轨迹的长度; RollCurve 在球为曲面接触时运动轨迹的长度; RelAngle 两个零件之间的相对转角; RelTrans 两个零件之间的相对位移; Screw 螺纹连接时的相对转角或相对位移。

f. Point 零件在某一点与基础连接。

g. Position 零件在某一点的坐标保持不变。

h. Relative 两个运动副运动规律之间的相互关系。

6) Joint 连接零件的关节

Joint 连接零件的关节有: Cylin 圆柱副; Revolute 转动副; Sphericl 球面副; Translat 移动副; Gear 齿轮副; Planar 平面运动副; PtCurve 点为曲线运动副; PtSurf 点为曲面运动副; SlideCrv 曲面滑动副; RollCrv 曲面滚动副; Bracket 接插连接 (自由度为零); Screw 螺纹连接; Universl 十字方向连接; 复合连接, 有 RevCyl、RevRev、RevSpher、RevTrans、Sphr-Sphr (这些复合连接在简化 DADS 模型时要用到)。

7) Force 产生力的单元

a. RSDA 扭转弹簧 (可考虑阻尼) 及在转动副上加扭转力矩。

b. TSDA 拉伸和压缩弹簧 (可考虑阻尼) 及在两连接点上施加作用力。

c. Friction 摩擦力单元 (定义各个运动副接触表面上的摩擦系数)。

d. 3Point 在零件上施加外力 (第一点定义外力作用点, 第二点及第三点定义力的作用方向)。

e. Beam 梁单元 (形状规则的零件, 考虑其弹性变形)。

f. Bushing 轴套单元 (考虑零件六个自由度方向的刚度)。

g. Leaf 汽车用板弹簧。

8) Other 辅助功能

a. Expression 按 C 语言的格式输入数学表达式 (得到的数值用了其他单元)。

b. Vehicle 将输出仿真得到的车辆运动状态 (在构造车辆的悬挂系统时将以此作为参照系)。

c. Suspension 建立车辆悬挂系统的模型: Double- $\Lambda$ -Arm 悬挂系统或称作 Short-Long-Arm 悬挂系统; MacPherson 悬挂系统; Five-Link 悬挂系统; Three-Link 悬挂系统; Iax-Roll 悬挂系统; Iax-Steer 悬挂系统; Hotchkiss 悬挂系统。

d. Curve 构造建模时要用到的各种曲线、曲面: SplineCurve 样条曲线; SplineSurf 样条曲面; ParmCurve 参数曲线; ParmSurf 参数曲面; LineArc 直线为圆弧复合曲线和曲面。

e. InitCond 给定在仿真开始时的各种初始值: One-Body 零件的绝对初始位置和初始速度; Two-Body 两个零件之间的相对初始位置和初始速度; Orient 零件的初始角速度; RelAngle 两个零件之间的相对角度和角速度; ExtraCrd 给定特别坐标系的初始位置和初始速度; ModeShape 给定零件某阶模态的变形及变形速度的放大比例。

f. PtIntrest 指定感兴趣的点的位置。用于仿真后输出动态响应值。

g. Stop 指定暂时停止仿真计算的时间。

h. Restart 指定从哪一个仿真计算的停止点开始继续进行仿真。

### (2) DADS Solver

DADS Solver 模块的功能是将在 DADS 中建立的模型进行装配分析、自动生成动力学方程、自动解算这些方程并给出结果。分析的类型有: Assembly (装配分析)、Static (静力学分析)、Kinematic (运动学分析)、Dynamic (动力学分析)、Invers...Dynamic (逆动力学分析)、Preload (预载荷分析)。在解算时能使用的积分器类型有: VARIABLE、PECE、RK、BDF。

### (3) DADSGraph Postprocessor

DADSGraph Postprocessor 模块, 简称 DADSGraph。用来显示及输出仿真后得到的系统运动学及动力学动态响应的数值及曲线。所有这些量都可以按全局坐标系及局部坐标系(零件质心处的坐标系)输出。包括:

1) 所有零件上任意点的位置、速度、加速度、角速度、角加速度以及在欧拉坐标系下的欧拉系数及其一阶二阶导数。

2) 所有运动副、零件之间连接点处的相互作用力及其幅值。

3) 所有 Force 单元(如弹簧、空气阻力、轮胎、摩擦力等)的计算值及其幅值。

4) 底座与基础之间的作用力及其幅值。

5) 驱动元件(如 Driver、扭转弹簧等)的驱动力矩及其幅值。

6) 柔性体的模态及每一个节点的位置、速度、加速度。

7) 液压系统中各元件节点(输入及输给点)上的压力、流量值。

8) 控制系统中各元件节点上的信号量。

9) 各种输入曲线。

10) 总能量曲线。

11) 用 C 语言表达式将这些量组合并绘出曲线(用于一些特殊的曲线)。

对这些曲线进行操作的菜单有:

1) Write 输出文件的格式有: ASCII、Curves (DADS 格式)、IDEAS、ARIES、ADAMS、Linear (线性化方程的系数矩阵)、PATran、Parameters (零件位置的数据及欧拉系数)、PRW (履带车超级模块中的零件位置的数据及欧拉系数)。

2) History 将当前绘制的所有曲线保存起来。

3) Operations 中的菜单有: Delete (删除)、Add

(加)、Subtract (减)、Mutiply (乘)、Devide (除)、Differential (微分)、Integrate (积分)、Scale (按比例放大或缩小)、FFT (快速傅立叶变换)

4) Appearance 确定曲线的外观, 如: 标题、横坐标及纵坐标的名称、坐标轴的颜色及曲线的颜色、曲线的线型、是否用对数坐标、是否画出网格、是否自动确定比例、给出横坐标及纵坐标的数值范围和单位长度。

5) Show 显示选定曲线的数值。

6) 所有的曲线及数据都能输入到文件或传输到打印机。

## 5.2.2 标准组合 DADS/Standard

基本组合 DADS/Basic + Control & Hydraulic Elements + User-Defined Forces & Contact Elements, 这种组合称为基本组合 DADS/standard。

### (1) Control & Hydraulic Elements

1) 通过在 DADS 中建立控制系统部分的模型: Amp 放大器; DeadZone 死区; Delay 延迟; 1stOrdA 一阶模拟控制器; 2ndOrdA 二阶模拟控制器; 1stOrdD 一阶数字控制器; 2ndOrdD 二阶数字控制器; Function 函数; Hystsis 滞后; Input 将零件上某一点的动态响应值作为控制系统的输入; Integrtr 积分; Limiter 限值; Multiplr 数值乘; Output 将控制系统信号转换成为力值或力矩值作用在机械系统上; PlantIn 在联合仿真时指定某个量从控制系统仿真软件中输入到 DADS; PlantOut 在联合仿真时指定某个量输出到控制系统仿真软件中; S Delay 延迟-一个取样周期; Summer 数值加; Switch 开关。

2) 通过 Hydraulic 模块在 DADS 中建立液压系统部分的模型: Accum 蓄能器; ChkValv 单向阀; SingAct 单作用缸; Dbl Act 双作用缸; Motor 液压马达; Valve 二位三通及三位四通换向阀; Servo 伺服阀; Line 管道; Orifice 阻尼孔; Parameter 液体的压缩率; Flow Vol 液体的流量; Volume 液体的体积。

### (2) User-Defined Forces & Contact Elements

1) Contact Elements。通过 Contact Elements 模块在 DADS 中定义接触和碰撞(指两个零件可以处在接触状态及分离状态)。只要选定接触面, 在“弹性模量”、“碰撞系数”和“摩擦系数”各栏中添入适当的数值即可。以上三栏也可以用“弹性系数”、“阻尼系数”或相应的曲线代替。

PointPoint 接触面是两个球面; PointSeg 一个球面与一个曲面接触; PointSpring 一个球面与一个旋转曲面接触; PointSegRail 轮轨接触; CAD 由机械 CAD 软件 CATIA 生成的曲面之间的接触。

2) User-Defined Forces。通过 User-Defined

Forces 模块用户可以在 DADS 环境中进行二次开发。在仿真计算时,各种力可以由三维机械模型以及仿真模型中的 Force 元素决定,用户也可以通过 User-Defined Forces 决定这些力。在这一模块中,软件提供了计算所有这些力的程序样板(Fortran 程序),用户只需要将计算这些力的代数表达式写入到相应程序的适当位置,DADS 就可以将这些“特殊”的力与仿真模型融为一体。所有样板程序中的全部变量都有详细说明。利用 User-Defined Forces 模块,可以建立复杂的仿真模型,可以便仿真模型更接近真实的实物样机。灵活应用这一模块,可以将所有的机械系统建成仿真模型。

### 5.2.3 高级组合 DADS/Advanced

标准组合 DADS/Standard + DADS/Flex, 这种组合称为高级组合 DADS/Advanced。

#### DADS/Flex

通过 DADS/Flex 模块,DADS 可以将零件的有限元分析结果中的模态读入,在仿真结果中将计算零件的“小变形”对机械系统动态响应的影响,也可以说,这种仿真模型中的零件是真实的状态,被称作“柔性体”。进行“柔性体”仿真有以下几个特点:

- 1) 计算得到的机械系统动态响应值更精确(包括各个零件上的受力)。
- 2) 为零件的应力分析、疲劳分析、设计优化提供更精确的边界条件(输入值)。
- 3) 动态显示零件在运动时的真实变形(可以放大显示)。
- 4) 预测机械系统的整体频率响应特性。

DADS 可以读入以下有限元软件的模态数据文件格式: NASTRAN; ANSYS; I-DEAS Master Series; ABAQUS; PolyFEM; CATIA 中的 Elfini。

### 5.2.4 专业模块及其他功能模块

#### (1) DADS/Pro、DADS/IMS 和 CATDADS

Pro/Engineer to DADS 模块可以嵌入到机械 CAD 软件 Pro/E 中,用户可以在 Pro/E 的环境中使用 DADS,在 Pro/E 的三维机械 CAD 模型基础上直接生成仿真模型;也可将 Pro/E 的模型输出到 DADS 中。

DADS/IMS 和 CATDADS 则分别是 DADS 和 I-DEAS Master Series、CATIA 的接口模块。

#### (2) DADS/Plant

DADS/Plant 由 DADS/Model、GUI、Animation 模块、DADSGraph Postprocessor 模块、SystemBuild Solver 模块、Simulink Solver 模块及 Easy5 Solver 模块组成。DADS/Plant 可以将机械系统的仿真模型与控制系统的仿真模型结合起来,进行联合仿真。用户不

必建立机械系统的数学方程,就能很快地生成机械系统和控制系统联合的整体仿真模型。

1) SystemBuild Solver 模块可以将 DADS 中的机械系统仿真模型与控制系统仿真软件 MATRIXx 的 SystemBuild 模块中的仿真模型联合起来。

2) Simulink Solver 模块可以将 DADS 中的机械系统仿真模型与控制系统仿真软件 MATLAB 的 Simulink 模块中的仿真模型联合起来。

3) Easy5 Solver 模块可以将 DADS 中的机械系统仿真模型与控制系统仿真软件 Easy5 中的仿真模型联合起来。

#### (3) DADS/Linearization and Eigenvectors

该模块可以将代表机械系统的非线性方程线性化,以一定的时间步长输出一系列线性方程。因为大多数的控制器设计是建立在线性控制理论基础上的,因此在分析机械系统的控制稳定性时,这些线性方程非常有用。该模块按一定格式输出的文件,可以被 MATRIXx、MATLAB 及 Easy5 读入。

#### (4) Advanced Tire

Advanced Tire 模块可以仿真出各种轮胎与地面的接触状态,计算出轮胎与地面之间的各种接触力及阻力矩等,这些力及力矩会对车辆的动态响应有很大影响。Advanced Tire 模块中有标准的轮胎模型、建立在 Pacejka 的“Magic formula”理论基础上的模型以及根据用户的方程定义的轮胎模型。

#### (5) Track Superelement

该模块是集成化的生成履带车辆与地面之间相互作用的超级模块。用户只须输入履带传动及地面的特征值,就可以定义好履带与地面之间相互作用的仿真模型。与用机械 CAD 模型生成履带传动仿真模型相比,该模块具有操作少、生成模型快、解算快、计算精度高等特点。综合效率至少提高 200 倍。

#### (6) Shock Absorber

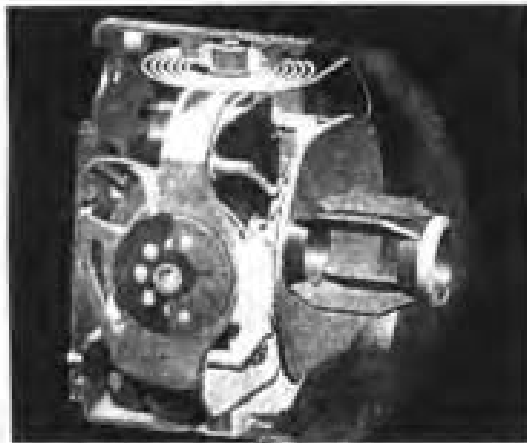
Shock Absorber 模块用来定义车辆上的减振器,仿真减振器的非线性阻尼,计算减振器两端的作用力变化及动态响应。适合用于中低频率的减振,最高振动频率为 30Hz,如用于这一频率以上的减振器可用弹簧—阻尼系统(TSDA)代替。

#### (7) DADS/Engine

在 DADS/Engine 中集成了一些常用于生成发动机仿真模型的模块。对于非发动机模型,如果结构类似,也可以用这些模块。其中 Hydrodynamic Bearing 模块用来定义动压油膜轴承的仿真模型,计算轴承内的接触力及摩擦力矩。此模块考虑了轴承内温度的变化对润滑油粘度的影响。HeliSpring 模块用来定义螺旋弹簧。与 TSDA(弹簧—阻尼系统)不同的是,

HeliSpring 模块考虑了弹簧的质量、转动惯量和弹簧端部对动态响应的影响。在弹簧低速运动场合,用 HeliSpring 模块和用 TSDA 定义弹簧区别不大。但如弹簧运动速度比较高(如发动机气门处的弹簧),用 HeliSpring 模块定义弹簧能保证仿真结果的精确度。Combustion Force 模块用来计算发动机活塞和气缸之间的作用力。用户给出发动机缸体气体燃烧爆炸时的气体压力和曲柄的转速、转角之间的关系曲线。该模块会计算出发动机活塞和气缸之间的作用力大小。Cam Contact 模块用来计算凸轮与滑杆之间的作用力。用 cam Generator 构造出凸轮的外形,给定凸轮与滑杆之间接触的刚度系数、阻尼系数及摩擦系数,Cam Contact 模块能计算出凸轮与滑杆在运动时的接触力、摩擦力及摩擦力矩的变化。

Lockheed Martin 公司开发机载激光吊仓产品见图附-1-1。



图附-1-1 机载激光吊仓产品

## 6 MSC. DYTRAN 软件

### 6.1 MSC. DYTRAN 概述

MSC. DYTRAN 诞生于 1991 年,是一个用于分析高度非线性、瞬态动力响应的大型通用三维有限元程序。美国 MSC 公司融合了它的两个产品 MSC/DYNA 和 MSC/PISCES 的精华于这一新系统。MSC/DYNA 是 MSC 公司对 LS-DYNA3D 进行全新改进后于 1988 年发布的,它具备优秀的显式积分算法和多种材料模式而被广泛用于分析大量的非线性瞬态问题,如高速撞击、爆炸、接触摩擦等。MSC 公司在 PISCES 欧拉模式算法的基础上,自行开发了物质流动算法和流固耦合算法,使得 MSC. DYTRAN 成为世界上第一个能够用完全耦合的方法解决流体-结构相互作用的问题,目前已经被广泛地应用于航空、航天、汽车、铁路、国防、核工业等领域。此外, MSC.

DYTRAN 在单元库、数据结构、前后处理等方面与 MSC/NASTRAN 取得了一致。

1994 年 MSC 公司收购了 PDA Engineering 公司后,开发了 PATRAN 与 DYTRAN 的界面接口。全新的、方便快捷的 MSC/PATRAN 前后处理器,使得 MSC. DYTRAN 向 CAE 自动化迈进了一大步。

### 6.2 MSC. DYTRAN 的算法特点

#### 6.2.1 MSC. DYTRAN 的拉格朗日方法(Lagrange 方法)和欧拉方法(Euler)方法

MSC. DYTRAN 具有两种求解方法即基于有限元法的拉格朗日方法和基于有限体积法的欧拉方法。这两种方法可以单独使用也可以一起使用,并可以将对应的拉格朗日网格和欧拉网格耦合在一起用来处理流固相互作用问题。当使用 Lagrange 方法时,单元的质量是不变的,节点在物体上的位置是固定不动的,物体单元由节点连接构成,物体变形时,节点物体在空间移动,单元亦发生相应的变形。Lagrange 方法是用于解决恒定质量单元即 Lagrange 单元的变形问题。这种方法适用于结构的力学分析,并为大多数有限元程序所采用。

当使用 Euler 方法时,单元的体积是不变的,节点在空间上是固定不动的,单元是由节点连接构成的,欧拉网格是一个固定的参考框架,物体的材料在分析过程中可以在网格中流动,并且材料的质量、动量和能量也随之从一个单元流到另一个单元。因此在划分单元时,欧拉网格的范围必须包含材料的所有可能流动区域。这种方法适用于分析气体、流体的流动和固体结构的极大大变形问题, MSC. DYTRAN 功能强大的欧拉法使其避免了三维网格重划分而能精确模拟大变形和极大大变形问题。

#### 6.2.2 显式时间积分方法

MSC. DYTRAN 采用了显式时间方法求解动力响应方程,这种方法因不需要对刚度矩阵进行分解而具有较高的解题效率和较低的存储要求。

与隐式时间积分法相比,显式法要求的积分时间步长要小得多而影响计算效率,而隐式法则要对刚度矩阵进行多次分解亦影响计算效率。在对于分析高度非线性动态问题,为了保证收敛性,无论用何种方法,积分时间步长都必须取得很小。由此可见,对于解决瞬态高度非线性问题,采用显式方法较隐式法要优越得多。

MSC. DYTRAN 的解题的有效性和计算精度已被许多考题和实际应用实例所测试和验证,是一种非

常优秀的非线性动态解题方法, MSC. DYTRAN 采用这种方法解决了大量隐式法无法解决的问题, 例如冲压成形中的皱褶、轮-轨接触、结构高速撞击、核管道甩动等问题。

### 6.2.3 有限体积方法

MSC. DYTRAN 的欧拉算法采用了有限体积法模拟材料的流动, 考虑了在有限体积内的质量守恒、动量守恒、能量守恒、状态方程和本构关系, 其功能十分强大, 可有效地解决大变形和极度大变形问题, 如: 液压传动分析、液体晃动、爆炸分析、高速穿甲和三维锻造成型等问题。

## 6.3 MSC. DYTRAN 的主要分析功能及其特点

MSC. DYTRAN 是一个有着通用目的和广泛用途的非线性动态三维有限元分析软件包。它采用高效的显式时间积分技术, 能模拟多种材料和几何的高度非线性问题。尤其在模拟高速碰撞、结构和零部件的大变形、液气体流动、流固耦合等方面, MSC. DYTRAN 的能力是独一无二的。

MSC. DYTRAN 采用拉格朗日和欧拉两种处理器对结构和流体进行建模。实体、壳、梁、薄膜、弹簧和刚性单元被拉格朗日处理器用于结构的建模; 欧拉网格由三维六面体单元组成, 欧拉处理器可以处理具有抗剪强度的材料的流动。

在材料模式中, 它包括了线弹性材料、弹塑性材料、刚性材料、橡胶材料、低密度泡沫材料、土壤材料、正交各向异性材料、层合复合材料、率相关材料以及各种屈服准则、失效模式、状态方程、爆炸燃烧模型等。

载荷模式可以是与时间有关的定向的或随动的集中载荷、压力以及各种初始条件。在欧拉网格中还可以定义具有各种性质的流动界面等。

约束有普通的节点约束 (SPC)、刚性节点约束、刚性墙约束、节点间的可分开连接 (BJOIN)、壳单元和实体单元节点连接 (KJOIN) 以及刚体间的连接。

接触问题包括考虑摩擦的面-面接触、点-面接触、单一平面自身接触、自适应(侵蚀)接触以及接触滑动等。MSC. DYTRAN 强大的接触算法功能使得可以很方便地分析物体跌落实验 (Drop Test) 仿真问题。

MSC. DYTRAN 在安全防护方面除了安全气囊的膨胀分析, 还与美国航天医学研究中心开发的 ATB 程序提供的假人模型 (此模型主要应用于动态运动中人体生物力学研究) 进行了全面集成, 使得 MSC. DYTRAN 可以在分析汽车碰撞时考虑人体的响应及运动姿态。

MSC. DYTRAN 的重启动功能可以十分方便地将一个规模较大的逐步分阶段进行运算。

### (1) 单元库

MSC. DYTRAN 具备一维、二维、三维单元库, 具体如下:

- 1) 杆单元 (CROD)。
- 2) 梁单元 (CBEAM, CBAR), 如 Belytschko-Schwer 模式, Hughes-Liu 模式, 任意梁截面库。
- 3) 安全带单元。
- 4) 点焊单元。
- 5) 四边形壳单元 (CQUAD4), 如 Belytschko-Tsai 模式, Hughes-Liu 模式, Key-Hoff 模式, 薄膜单元。
- 6) 三角形壳单元 (CTRIA3), CO 模式, 薄膜单元。
- 7) 体单元 (CHEXA, CPENTA, CTETRA)。
- 8) 弹簧/阻尼单元 (CELAS, CDAMP), 如线性单元、非线性单元、扭簧, 用户自定义单元 (子程序 EXSPR, EXVISC, EXELAS)。
- 9) 集中质量单元 (CONM2)。
- 10) 刚性单元, 如椭球体 (RELLIP)、刚性体 (RIGID)、刚性单元 (RBE2)、刚性墙 (WALL)。
- 11) 欧拉单元 (CHEXA, CPENTA, CTETRA), 如单材料实体单元和多材料实体单元。

### (2) 材料库

MSC. DYTRAN 采用组合方式利用十四张材料卡片可定义近百种材料模式, 另外, 可通过用户子程序定义各种破坏模式。

1) DMT 卡。为通用的材料本构关系, 可通过设置其中的各个参数进行组合定义具有以下任一性质的材料:

- a. 剪切模式: 常剪切模式 (SHREL); 线性粘弹性剪切模量 (SHRLVE)。
- b. 屈服模式: 流体动力屈服模式 (YLDHY); Von Mises 屈服模式 (YLDVM); Johnson-Cook 屈服模式 (YLDJC)。
- c. 状态方程: 理想气体状态方程 (EOSGAM); 多项式状态方程 (EOSPOL); 体积模量状态方程 (EOSPOL); Tait 状态方程 (EOSTAIT); JWL 状态方程 (WOSJWL); 用户自定义状态方程 (EXEOS)。
- d. 失效模式: 最大塑性应变失效模式 (FAILMPS); 用户自定义失效模式 (FAILEX, FAILEX1, EXYLD); 最大等效应力和最小时间步长失效模式 (FAILEST); 最大等效应力失效模式 (FAILMES); 最大应力失效模式 (FAILPRS); 最大塑性应变和最小时间步长失效模式 (FAILSDT)。
- f. 层裂模式: 最小压力层裂 (PMINC)。

- 2) DMATEL 卡——弹性材料。
- 3) DMATEP 卡——考虑失效的弹塑性材料。
  - a. Von Mises 屈服模式 (YLDVM)。
  - b. Johnson Cook 屈服模式 (YLDJC)。
  - c. 最大塑性应变失效模式 (FAILMPS)。
  - d. 用户自定义失效模式 (FAILEX)。
- 4) DMATOR 卡——正交各向异性弹性材料。
- 5) MAT1 卡——考虑失效的纤维复合材料。
  - a. Tsai-Hill 理论。
  - b. Tsai-Wu 理论。
  - c. 修正的 Tsai-Wu 理论。
  - d. 最大应力理论。
  - e. Hashin 理论。
  - f. Chang-Chang 理论。
  - g. 用户自定义 (EXCOMP)。
- 6) SHEETMAT 卡——非均质塑性材料，主要用于钣金成形。
  - a. Krieg 屈服模式。
  - b. 塑性硬化。
  - c. 应变率相关。
- 7) DYMAT14 卡——土壤和可压泡沫材料。
- 8) DYMAT24 卡——分段线性塑性材料。
- 9) DYMAT26 卡——可压正交各向异性材料。
- 10) MATRIG 卡——刚性材料。
- 11) RUBBER1 卡——橡胶材料。
- 12) FORM1 卡——泡沫材料 (聚丙烯材料)。
- 13) FORM2 卡——泡沫材料；用户可定义应力应变曲线；率相关。
- 14) 用户子程序：EXFAIL 失效模式；EXLAIL1 正交各向异性实体定义失效模式；EXCOMP 复合材料失效模式；EXYLD 屈服模式；EXBRK 可断开连接的失效模式；EXEOS 状态方程。

### (3) 接触解法

- 1) 主从面接触：变形体-变形体接触；变形体-刚性体接触。
- 2) 主从点接触：节点-变形体接触；节点-刚性体接触。
- 3) 单面自身接触。
- 4) 自适应接触 (侵蚀接触)。
- 5) 库伦摩擦，静、动力摩擦系数和滞留系数。
- 6) 考虑壳单元厚度及间隙。
- 7) 粘合连接：面-面连接；点-面连接；线-面连接。

### (4) 刚性体

- 1) 用解析法定义的刚性椭球。
- 2) 任意形状的刚性体：用刚性面定义；用刚性材料定义。

### (5) 边界条件和约束

- 1) 单点约束 (SPC)。
- 2) 刚性网格点连接 (RBE2)。
- 3) 刚性墙。
- 4) 可断开连接 (BJOIN)。
- 5) 壳体与实体间的动态连接 (KJOIN)。

### (6) 载荷

- 1) 集中力和集中弯矩。
- 2) 跟随力和跟随弯矩。
- 3) 压力载荷。
- 4) 重力载荷。
- 5) 强迫速度。
- 6) 初始条件。
- 7) 气袋压力。

### (7) 用户自定义程序

- 1) 定义材料的失效。
- 2) 定义压力载荷。
- 3) 定义状态方程。
- 4) 定义弹簧/阻尼特性。
- 5) 定义速度场。
- 6) 用户指定的节点和单元的输出请求。

### (8) 流体~结构的耦合解法

通常的软件在处理流体-结构相互作用问题时，将流体产生的力作为“预先确定”的载荷作用到结构上进行分析。而 MSC. DYTRAN 则不然，它是通过直接耦合结构网格 (Lagrange 网格) 和流体材料网格 (Euler 网格) 间的响应自动地、精确地算出流固界面处的物理性质。在这个过程中，Euler 材料流动引起的压力载荷通过耦合算法自动作用到结构的有限元网格上。在这种压力作用下，结构的有限元网格将发生变形，结构的变形也反过来影响 Euler 材料的流动和压力值。这种结构变形和流体载荷间的相互影响使得我们可以得到完全耦合的流体-结构响应。

MSC. DYTRAN 具有三种处理 Lagrange 网格和 Euler 网格耦合分析方法：

1) 普通耦合法 (General Coupling)，适用于所有的耦合问题，如安全气囊的膨胀过程、子弹穿甲过程、炸弹爆炸分析等。图附.4-5 为汽车-气囊-人体系统仿真分析。

2) 快速耦合法 (Fast General Coupling)，要求 Euler 网格法向与总体坐标系平行，此方法大大加快耦合算法。

3) ALE 法 (Arbitrary Lagrange Euler)。用 ALE 法，Euler 网格可以随着结构的变形而移动，它是一种求解效率较高的方法，这种方法的典型应用是鸟体碰撞飞机结构的过程、液体晃动等分析。

## 6.4 MSC. DYTRAN 的主要特色

1) 最适合非线性动态或非线性准静态问题的分析, 可求解高速瞬态大变形分析。

2) 友好的用户界面, 工业标准的 MSC/NASTRAN 格式的输入数据文件, 因此与 MSC/NASTRAN 作接口的前后处理器均可与之作接口。

3) 与功能强大的 MSC/PATRAN 集成的前后处

理界面, 可以使 DYTRAN 的使用变得更简单, 更直观, 更易学。

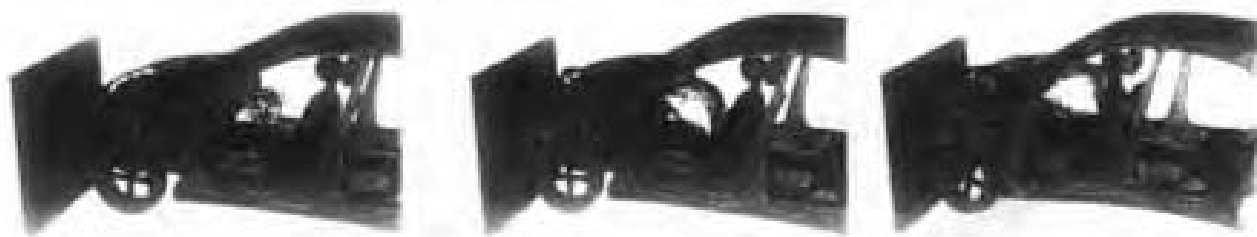
4) 具有非线性材料模型的完整的单元库。

5) 功能强大, 考虑摩擦效应的三维接触算法。

6) 用单一的分析软件包可完成有限元结构分析、材料流动分析、流体结构耦合分析等多种分析类型。

7) 开放式接口, 用户可自定义多个子程序。

8) 在许多工业应用中被证实了的可靠运算技术。



图附-1-3 汽车—气囊—人体系统仿真分析

## 7 铸造工艺分析软件 InteCAST 系统

InteCAST——铸造工艺分析软件系统是华中理工大学开发的, 用来分析和优化铸件铸造工艺的重要工具。它以铸件充型过程、凝固过程数值模拟技术为核心对铸件进行铸造工艺分析, 可以完成多种合金材质 (包括铸钢、球铁、灰铁、铸造铝合金等) 和多种铸造方法 (砂型铸造、金属型铸造、低压铸造、铁模覆砂铸造等) 下铸件的凝固分析、流动分析以及流动和温度的耦合计算分析, 曾在多种不同材质复杂铸件的工艺改进、工艺优化中圆满地完成增收降耗的任务。InteCAST 在预测铸件缩孔缩松的倾向, 改进和优化工艺, 提高产品质量, 降低废品率, 减少浇冒口消耗, 提高工艺出品率, 缩短产品试制周期, 降低生产成本, 减少工艺设计对经验人员的依赖, 保持工艺设计水平稳定等诸多方面都有明显的效果。

### 7.1 主要功能模块

#### (1) 前置处理模块

铸件形成过程中一些重要的物质过程, 如流动过程、传热过程、凝固过程等, 都可以用一定的数学方程, 即偏微分方程来描述, 在复杂的边界条件下, 这类方程一般不存在解析形式的解, 但计算机的高速度大容量使得我们可以以任意逼近的近似程度求得数值解。其原理是将连续的三维场变量离散成时空坐标上有限的分离变量, 将描述连续场的偏微分方程演绎成有限单元场的差分方程, 通过大量的迭代计算进行求解, 前置处理模块的任务就是为铸件实体的迭代计算准备一个离散的三维场空间, 这一过程俗称网格划分。

本模块运行于 AutoCAD R12.0 for Windows 平

台, 借助其有关功能完成铸件的图形输入及几何造型, 并通过简单的自动加载, 调用网格剖分子模块, 对铸件三维造型实体进行网格划分, 为后续计算模块提供原始的网格数据 MAT 文件。该模块的特点有:

1) 铸件图形输入以 AutoCAD 为操作平台, 用户基础好, 易于推广。

2) 可处理多物性系统 (包括铸件、砂型、铁模、砂芯、冒口、冷铁等十多种物性系统)。

3) 全部操作都用鼠标完成, 使用方便快捷。

4) 坐标空间、剖分方案、网格尺寸、搜索顺序等都可选可调, 功能实用, 操作灵活。

5) 网格自动生成, 省时省心。

#### (2) 计算处理模块

计算分析模块是系统的核心, 本软件系列包含铸钢、球铁、灰铁和铸造铝合金等类材质铸件的凝固分析和流动分析等多种可选模块, 可完成相应材质铸件在砂型、金属型、低压、铁模覆砂等铸造方式下的温度场计算、流动场计算和流动场温度场耦合计算。在低压铸造模块中, 还提供升液过程模拟, 多周期浇注模具温度分析, 模具预热、冷却、保压等参数优化以及模温过热度预测等功能, 为实际生产提供全面的科学分析。

计算模式有两种, 一种是单纯温度场的计算, 它忽略充型过程, 假定均一的初始温度, 因而有一定的偏差, 但可显著地缩短计算时间, 如果计入充型过程对初始温度的影响, 就要进行流动、温度的耦合计算, 这是另一种计算模式。后一种计算模式精度高, 但计算量大, 比较费时, 一般在充型过程不长、对初温影响不大的情况下, 可不必用此模式。该模块的特点有:

1) 凝固分析以固相率为准则, 处理单元数超过

200 万。

2) 微机上实现实用铸件流动场的计算,可处理流动域单元数超过 30 万。

3) 多任务界面,交互式操作,运行状态,运行参数随时可调。

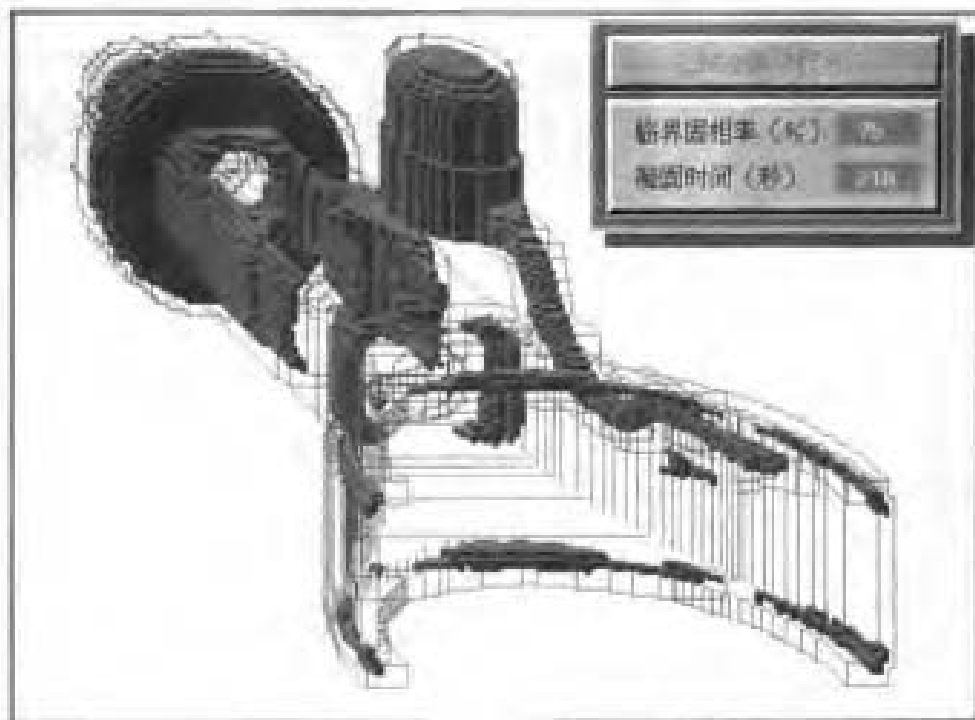
4) 完整保存、完全恢复计算间断处的现场状态,以便随时停机开机,保证计算的连续性。

5) 低压铸造模块,除一般的凝固,流动分析外,还

提供升液过程仿真,多浇注周期模温分析,模具参数优化,保压、冷却工艺优化等实用性很强的功能。

(3) 后置处理模块

通过大量迭代计算得到的许多数据,包含了凝固过程丰富的信息。准确如实地以恰当的方式将这些信息传达给用户,是最后完成整个软件分析功能极其重要的一步。本模块包括曲线生成,图形生成,动画制作,动画演示等功能。



图附 4-4 铸件凝固过程动画之一帧

对于铸件的凝固分析,其结果显示分凝固过程显示和温度过程显示两种方式。凝固过程显示方式只着色显示未凝固的网格单元,而隐去已凝固单元,使其透明。同时用细线条勾勒出铸件型腔的线框图形。这样就可直观透彻地观察铸件的凝固过程,观察冒口的补缩效果。温度过程显示方式是根据色温对应关系着色铸件的各单元,显示铸件的温度分布,画面层次丰富细腻,效果逼真。图附 4-5 为铸件凝固过程动画。

流动分析结果的表达方式有三种,体积、流速、流线。体积方式是只着色已充满流体的单元,可以明确地显示流体的充盈状态。流速是流场的矢量表达方式,用带箭头的线段显示,其长度表示流速大小,其指向表示流动方向,每单元对应一个流速矢量,可以显示速度分布,也可显示充型状态。流线则是流速的积分形式,是速度矢量的包络线。图附 4-7 为流速矢量的三维分布。

但是,仅仅用静态画而来表达一个复杂过程,是不完整的,耗费大量机时计算出来的数据,其大部分信息

隐含在过程的变化中,这些变化的动态的信息若不充分利用,显然是一种极大的资源浪费。本软件除了可显示各种静态画面外,还按照流行的 FLC、AVI 动画格式,提供动态演示的各种功能,包括画面剪辑,动画制作,播放控制,动画演播等。用动画显示凝固过程,充分地利用了数据的信息资源,不遗漏任何不该忽略的细节,可使用户完整无遗地观察整个过程的变化。

## 7.2 主要特点

1) 铸件流动场,温度场计算结果多种形式的二维、三维静态显示。

2) FLC 动画、AVI 动画编辑制作系统界面友好,操作方便,自动压缩,自动合成。

3) 通过三维流动数值模拟动画,能够观察分析铸件浇注系统的流束导入、速度分布,压力分布,流量分布以及卷气冲砂等情况,为改进优化浇注系统提供充分的依据。





图附-1-1 温度场数值分析

### 7.3 HSC 注射模流动保压冷却模拟分析系统

华中模具技术国家实验室推出的 HSC 2.5 系统,是采用先进的软件工程技术开发成功的新一代注射模 CAD/CAE/CAM 集成化系统。HSCAE 软件是一组在模具制造之前能预测塑料熔体在型腔内流动、保压、冷却行为的程序。采用这些分析程序,能保证模具设计质量,提高试模一次合格的成功率。

HSC2.5 是基于微机平台开发的注射模设计、分析与制造集成系统,可在各种操作系统环境下运行,面向广大模具设计人员,具有方便实用的中西文 WINDOWS 界面,完善的培训和技术服务。塑料注射模流动、保压及冷却模拟软件能够以交互方式预测塑料熔体的充满位置和熔合纹形态,使流动与保压实现了联机集成,能够提供熔体在模具内各时刻的压力场、温度场和密度场,优化模具的浇注系统和注射成型工艺参数。达到了 20 世纪 90 年代中期国外同类优秀软件的先进水平,被模具设计人员称为“试模软件”。对于减少模具返修报废,提高塑料制品质量具有显著的经济效益。经国内外数十家工业企业的验证与改进,可在机械、电子、化工、轻工、汽车等行业广泛推广应用。

主要功能有:

1) 注塑工艺条件初选:用户输入塑料制品的材料和基本尺寸后,可得到有关注塑填充时间、熔体温度、注射压力的推荐值。

2) 流变参数拟合:将流变仪测得的实验数据拟合为粘度模型中的流变参数。这些参数是运行流动模拟软件所必需的。

3) 流道平衡设计:当设计一模多腔的模具时,可利用该软件来计算各流道和浇口的合适尺寸,保证每股熔体能在等温等压下同时充满各型腔。

4) 流动与保压模拟:采用先进的有限元/有限差分耦合法,充分发挥了有限元法和有限差分法各自的优点,对复杂边界的适应性很强。在综合比较了各种数

值计算方法的优缺点的基础上采用有限元/有限差分耦合法求解流动仿真问题,并用控制体积法处理移动边界问题,取得了很好的效果。

5) 冷却分析:采用边界元法算法,可分析模具、制品及冷却管道在注塑过程中的冷却情况,既能提供稳态冷却结果,又能提供非稳态冷却结果。图附-4-5 为注射模冷却模拟。



图附-4-5 注射模冷却模拟

HSC 塑料注射模流动、保压及冷却模拟软件所采用的有限元/有限差分/控制体积算法,是求解三维流动仿真问题的一种行之有效的办法。采用有限元与有限差分耦合的方法,能为用户提供熔体在型腔内流动的动态图,提供不同时刻熔体各处的温度、压力、剪切速率、剪切应力以及所需的最大锁模力等,其预测结果对改进模具浇注系统的设计有着重要指导意义。

## 8 动力学分析设计师 Dynamic Designer/Motion

### 8.1 软件介绍

Dynamic Designer/Motion (简称 DDM) 是美国 MB 公司的产品。它是在功能强大、独树一帜的机械

系统动力学仿真分析软件——ADAMS 软件的基础上,为满足微机用户而专门开发的。用户如果需要,可以很方便地把 DDM 更新为 ADAMS 仿真软件包。

AutoCAD 和 MTD 是全世界最流行的计算机辅助设计系统。ADAMS 的动力学仿真分析模块被无缝地集成在 AutoCAD 和 MTD 之中。因此,用户可以借助于功能强大、已被广泛接受的 ADAMS 的动力学仿真功能,在 AutoCAD 或者 MTD 中,来演示和研究他所设计的复杂机构的运动规律。

DDM 加强了 AutoCAD 和 MTD 的功能。AutoCAD 和 MTD 借助于 DDM 的运动输入功能和约束输入功能,建立复杂机械系统的动力学仿真分析模型。通过求解该模型,用户能够对复杂的机械系统,进行三维动力学仿真分析研究。

用户如果拥有了 DDM,那么借助于上述仿真分析研究方法,他就可以在既没有切削任何一块金属,也没有铸造任何一块塑料之前,用计算机十分完善而精确地仿真分析出机械系统的运动规律。从而把从前需要数个星期乃至数月才能完成的物理模型和物理试验,缩短到仅要数个小时就能在计算机上完成。

## 8.2 DDM 的三维动力学仿真分析功能

DDM 可以帮助用户解决远比二维连杆机构更复杂的三维机械设计问题。而且更重要的是,用户可以直接在 AutoCAD 或者 MTD 中仿真分析和动画演示机构运动过程。

以汽车转向机构为例,如果使用 DDM,用户不但可以仿真该机构的运动情况,而且还可以计算出由于路面不平而引起的前轮摆动,传递到方向盘上产生的力的大小。

另外,DDM 可以对一个任意复杂的机器进行仿真分析,这些机构中可以包含力的影响因素,如弹簧、凸轮、日内瓦机构等。

问题的关键是,您所设计的对象是否运动?如果运动,那么就可以肯定,DDM 一定会更加快速地优化您的设计方案。

## 8.3 DDM 与 AutoCAD 和 MTD 界面相同,风格一致,数据共享

DDM 蕴含在 AutoCAD 和 MTD 之中。因此它采用 AutoCAD 扩展用户界面和 MCAD 编程的开发界面。DDM 在 AutoCAD 和 MTD 中起动并运行。起动之后,DDM 的界面与 AutoCAD 的界面相同,风格一致,数据共享,标准菜单一样,对话框也一样。倘若要从 AutoCAD 或者 MTD 中进入 DDM,只需在 AutoCAD 或者 MTD 的菜单和图标中选择一下即可。当 DDM 完

成仿真之后,它可以生成详细的零件图和部件图。而这些零件图和部件图,能够被全世界最流行的 CAD 软件所读写。

DDM 中,无论用户在哪个建模阶段,均能得到情景敏感帮助和热线练习。这使得 DDM 易学易用。大多数用户认为,他们仅需几个小时,就能在模型的运动中见到自己的设计思想。

## 8.4 DDM 的优点

1) 节省时间。利用 DDM,在产品开发的早期阶段发现设计中的问题,从而缩短产品开发时间。

2) 降低成本。利用 DDM,在产品开发过程中,降低对物理样机的依赖程度,从而降低产品开发成本。

3) 提高质量。利用 DDM,在产品开发的后期限阶段优化系统的整体性能。

4) 增加交流。利用 DDM,很容易在同事之间讨论设计思想,交流设计概念。

## 8.5 DDM 的建模器功能

1) DDM 的三维(和二维)建模器,方便易用,集成在 AutoCAD 和 MTD 中。

2) DDM 建模器利用所有类型的几何图形,例如线框图、着色曲面图、参数化实体和通过布尔运算所得到的组合形体图,作为机构中的零件。

3) 几何图形联想,约束联想,参数联想。

4) 在 DDM 的模型导航器中,选择欲在其上建约束的零件,只要拖拉鼠标就在零件之间自动建立起约束。

5) 如果零件是实体,则 DDM 自动计算该实体质量特性参数。

6) DDM 支持的约束或称运动副类型:转动副、圆柱副、平面副、齿轮齿条副、齿轮副、移动副、球面副、螺旋副、常速度约束、固定约束、万向节、槽约束、点线副、锥面平面副、锥面柱面副、球面平面副、球面锥而副、球面柱面副。

7) DDM 支持的随时间而变的力和力矩类型:拉压弹簧和扭转弹簧;拉压阻尼和扭转阻尼;作用力、作用力矩、反作用力、反作用力矩;碰撞力。

8) DDM 具有移动和转动运动激励生成器。

9) DDM 具有仿真凸轮及从动件和其他高副接触的功能:

a. 点线接触。两个零件在作相对运动时,一个零件上的某一点在另一个零件上某一曲线上运动。

b. 线线接触。两个零件在作相对运动时,一个零件上的某一曲线始终与另一个零件上某一曲线上相接触。

10) 定义、改变和使用不同的单位制。

11) 检查模型的完善性以便尽可能地减小求解误差。

## 8.6 DDM 求解器的功能

1) DDM 支持多零件系统的三维装配、静态装配、动力学分析、静态动力学分析等各种仿真方式。

2) DDM 支持基于目标的、先进的求解算法。使用户即便是分析极其复杂的机构系统,也能快速地获得仿真结果。

## 8.7 DDM 结果演示器的功能

1) DDM 用 X、Y、Z 三方向以及其的和的向量曲线图,显示任何一个零件上的任何一点处的位移、速度和加速度。

2) DDM 用 X、Y、Z 三方向以及其的和的向量曲线图,显示分离体上的惯性力、惯性力矩和受到的作用力、作用力矩。

3) DDM 用 X、Y、Z 三方向以及其的和的向量曲线图,显示其他任何通过计算才能获得的物理量。

4) DDM 用线框图、消隐线框图、着色实体图,动画演示系统的运动过程。

5) DDM 对系统中的每一个零件和弹簧,在其运动范围内进行干涉性检查。

6) DDM 显示任何一个零件上的任何一个点的运动轨迹。

7) DDM 根据需要,输出凸轮的轮廓曲线。

8) DDM 对结果曲线进行任意叠加。

9) DDM 保留各个设计方案的结果集。

10) DDM 中,用户只要选择一下菜单,就可以生成 AVI 文件。这些文件可以用 Microsoft Word 和 Microsoft Powerpoint 进行读写。这使得用户用 Microsoft Word 编写的仿真研究报告更加生动,用 Powerpoint 做的设计方案演示更加直观、形象并富有娱乐性。

## 8.8 DDM 的使用方法

(1) 在 MDT 中 如果用户使用的是 MDT,那么若要建立机械系统动力学仿真分析模型,只需在 DDM 的建模器中,把部件拖到零件“树”上某一节点上即完成建模工作。并且,自动创建零件间的约束、自动给出零件的质量特性参数。

(2) 在 AutoCAD 中 如果用户在 AutoCAD 中

绘制了零件,那么 DDM 利用这些零件,也能轻而易举地完成机械系统动力学仿真分析模型的建模工作。

首先,用户可以利用 AutoCAD 中的任何类型的几何图形,也包括线框图,直接作为三维 DDM 的机械系统动力学仿真分析模型中的连杆。一旦用户选定了将要作为连杆的几何图形,DDM 就会立即给出其质量特性参数。

其次,零件与零件之间的连接关系,通过定义运动副(或约束)来完成。DDM 支持的运动副(或约束)有 10 种之多。在定义运动副(或约束)时,只需给出零件之间的连接位置即可。如有必要,还可以定义高副约束,如点对曲线约束和曲线对曲线约束等。然后在机构上施加运动激励和力就完成机构的建模。

建好模型之后,DDM 自动检查 AutoCAD 和 MTD 中的所有零件的干涉性。只要用户选定某一待查零件,DDM 就会检查该零件与机构中其他零件在其运动范围内的干涉情况。一旦发现干涉时,DDM 就会高亮度显示干涉零件的外形。并且准备好一份干涉检查报告。

DDM 显示机构系统运动过程中,零件上的某一点的轨迹。这些轨迹可以用来确定机器的壳体形状和凸轮的外形轮廓。

DDM 可以显示零件上的任一点处的位移、速度和加速度。无论是在动画显示时,还是在 X? /FONT>Y 曲线图中,您可以用鼠标询问零件上的任一点处的运动信息。

DDM 可以在另外的窗口中,显示 X? /FONT>Y 曲线。该曲线可以与动画仿真一同显示。以增加视觉效果并帮助您理解设计方案。您还可以利用 Windows 的剪切板,把机构模型、图和仿真结果粘贴到任一类型的文件中。

## 8.9 软件环境

DDM 可以在 Windows95 和 Windows NT 4.0 平台上 AutoCAD 和 MTD 上运行。

### 参考文献

- 1 方明伦,俞涛,李莉敏. PDM 原理及应用. 上海:上海大学出版社,1998
- 2 张苏苹,尤嘉庆. AutoCAD 中文版升级教程. 北京:电子工业出版社,1999
- 3 樊炳辉,李云江. MDT——三维参数化特征造型机械 CAD. 北京:清华大学出版社,1999